INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
2025, VOL. 41, NO. 14, 9097–9108
https://doi.org/10.1080/10447318.2024.2423107

Predicting Engagement of Older People's Virtual Teams from Video Call Analysis

Nicoletta Noceti[a], Simone Campisi[a], Alice Chirico[b], Vittorio Cuculo[c], Giuliano Grossi[d], Monica Michelotto[b], Francesca Odone[a], Andrea Gaggioli[b,e]*, and Raffaella Lanzarotti[d]*

[a]MaLGa-DIBRIS, Università Degli Studi di Genova, Genova, Italy; [b]Research Center in Communication Psychology (PsiCom), Università Cattolica Del Sacro Cuore, Milano, Italy; [c]AlmageLab, Università Degli Studi di Modena e Reggio Emilia, Modena, Italy; [d]PHuSe Lab, Università Degli Studi di Milano, Milano, Italy; [e]IRCCS Istituto Auxologico Italiano, Milano, Italy

ABSTRACT

This study examines seniors' creative engagement in group activities using synchronous communication tools and explores automatic assessment methods through behavioral and psychophysiological measurements. Working with a small senior group on collaborative creative tasks, we implemented a comprehensive data collection approach using audio-visual and physiological measurements. Machine learning models were used to evaluate group creative engagement levels using various data subsets. Results show that engagement assessment can be effective with different feature combinations, allowing flexibility across contexts and constraints. The multimodal approach, combining facial, audio, and body analysis, achieved optimal performance and is recommended when conditions permit. Our research provides insights into seniors' online creative participation and presents an automated system for detecting creative engagement in virtual teams, supporting active participation strategies.

KEYWORDS
Engagement; virtual teams; audio-visual analysis; psychophysiological measurements; multimodal classifier

1. Introduction

The growing ageing population has brought increased attention to the cognitive decline associated with normal ageing, often leading to mild disabilities and potentially progressing to dementia. With limited success in developing effective drug treatments for dementia, there is a growing emphasis on lifestyle factors such as diet, exercise, and mental stimulation to maintain brain health in older adults (Phillips, 2017). Creative engagement stands out as a particularly beneficial activity, as numerous studies have shown its positive impact on cognitive abilities and overall brain health. This study explores the creative engagement of seniors in virtual teams using synchronous communication tools and considers behavioural and psychophysiological measurements for its automatic assessment. Our approach is grounded in the Networked Flow model (Gaggioli et al., 2013, 2015, 2020), which posits that collective creativity flourishes under conditions of high-level social presence and shared flow states. A multifaceted measurement approach, including audio-visual and physiological data, has been set up to collect data describing each individual in the group. Then, a group representation derived from the data is provided to a machine learning model to estimate the level of creative engagement of the group.

The paper is structured as follows. Section 2 provides the background and conceptual foundations of the study. Section 3 describes the experimental methodology. Section 4 details the procedures for acquiring, encoding, and processing multimodal data. Section 5 discusses the features used to develop the classification model for engagement values. Finally, the conclusion section offers a general discussion of the results, potential implications, limitations, and future research directions.

2. Background and contributions

2.1. Related works

There is growing evidence that participation in creative activities can foster a sense of control and promote positive social interactions, enabling older people to strengthen their resilience in meaningful ways. These benefits are rooted in the motivational, attentional, affective, and social dimensions of such activities (McFadden & Basting, 2010). In particular, the quality of human relationships, the emotional closeness, and the frequency of encounters are associated with positive health outcomes as they carry emotional significance. Social creative activities, including structured and unstructured leisure, are typically conducted in day centres or within geriatric hospitals' Residential Health Assistance facilities. However, numerous elderly people face obstacles to regularly participating in group recreational and social activities. Poor health, family facility staffing, transportation, and geography can influence the social interaction and participation of elderly people in these activities (Thomas et al., 2013).

CONTACT Vittorio Cuculo vittorio.cuculo@unimore.it Università degli Studi di Modena e Reggio Emilia, Modena, Italy
*Co-last author

Firstly, seniors may have limited mobility due to factors like frailty or neurological conditions, making it harder to access day centres. Additionally, the availability of these activities is often constrained in public geriatric facilities, while private facilities can be financially out of reach for many seniors. Weather conditions and other unforeseen events, such as the COVID-19 pandemic in 2020-2021, can also impose limitations on mobility and thus participation in relevant social activities.

Virtual teams have emerged as a potential solution to improve social interaction activities for the elderly (Walsh, 2019). In this study, we define virtual teams as groups of geographically dispersed individuals who work together using digital communication technologies to achieve a common goal (Lipnack & Stamps, 2000). However, there is a notable scarcity of empirical research in this area, particularly concerning older adults. A crucial research question revolves around how to effectively improve the level of engagement in online social interactions among this demographic (Burke et al., 2010; Hofer & Hargittai, 2024; Nguyen et al., 2022). Studies on creativity in virtual teams, although not specifically focused on older adults, provide valuable insights that may apply to our context. Chamakiotis et al. (2013) highlight the interplay between technology, teams, and individuals, highlighting that appropriate technological tools and effective team management can mitigate the challenges posed by geographical dispersion and asynchronous communication, thus improving creative results. Abi Saad and Agogué (2023) underscore the importance of adapting communication strategies and leadership styles to the unique needs of virtual teams to maintain high levels of engagement and creativity. Chamakiotis and Panteli (2023) examine temporary virtual project teams, suggesting that the temporary nature of such teams can either constrain or enhance creativity, depending on management and technological support. Ocker (2005) provides a qualitative analysis of asynchronous virtual teams, identifying key factors such as stimulating colleagues, a collaborative team environment, and effective management of technical difficulties as crucial to maintaining engagement and fostering creativity.

Furthermore, the impact of the use of online communication tools on the mental health of older people is mixed, with some studies suggesting a reduction in loneliness and an improvement in mental health, while others indicate potential negative effects, particularly when Information and Communication Technologies (ICTs) are used problematically (Meshi et al., 2020). Additionally, the use of online communication platforms such as Zoom or Meet during the COVID-19 pandemic has highlighted the utility of these tools (Menary et al., 2023) but also their potential limitations. For example, the so-called *Zoom fatigue* (Lim, 2023; Nadler, 2020) has become a common negative experience through the prolonged use of these platforms, which can decrease the level of engagement and social presence. This phenomenon may be induced by several dynamics, which include awkward turn-taking, inhibited spontaneity, restricted motility, lack of eye contact, and increased self-awareness (Aagaard, 2022).

2.2. The present study

Given these gaps in the literature, there is a clear need for research that specifically addresses the unique challenges and opportunities for fostering creativity and engagement in elderly virtual teams. Our study defines creative engagement through the lens of the Networked Flow model (Gaggioli et al., 2020, 2015, 2013), which conceptualises collective creativity as an emergent phenomenon arising from social interactions and shared experiences. This model builds on Csikszentmihalyi's concept of flow - an optimal psychological state characterised by intense focus, effortless action, and intrinsic motivation (Czikszentmihalyi, 1990; Mihaly, 1975). Networked Flow extends individual flow to collective flow, where group members experience a shared state of optimal engagement and creativity that can emerge and propagate through social networks, both in-person and online. Specifically, we define creative engagement as: (i) Active participation: The extent to which older adults contribute ideas, respond to others, and remain involved in the creative task; (ii) Cognitive involvement: The level of focus, attention, and mental effort directed towards the creative activity; (iii) Emotional investment: The degree of enthusiasm, interest, and positive affect displayed during the virtual interaction; (iv) Collaborative behavior: The willingness to build upon others' ideas, offer constructive feedback, and work towards shared goals; (v) Adaptability to the virtual environment: The ability to navigate and utilise the digital platform effectively for creative expression. Building on this foundation, our study seeks to operationalise and measure these dynamics in the specific context of virtual teams made up of older adults. Accordingly, we aim to address the following research question:

> RQ: How can we effectively measure and predict creative engagement in virtual teams of older adults using multimodal data and machine learning approaches?

As - to the best of our knowledge - this research is the first of its kind, a custom ad-hoc methodology was developed to study, measure, and classify creative engagement (i.e., Networked Flow) experiences of virtual elderly teams using a machine learning approach. The research protocol involved assigning participants to small groups, each moderated by a researcher. Each of these groups was then given a collaborative creative task (i.e., inventing a last-minute recipe with available cupboard ingredients) to be completed remotely.

Our operationalisation of creative engagement incorporated multiple measures to capture the complex nature of this construct:

1. Expert ratings: Two independent raters assessed the creative engagement of each participant using a Likert-type 5-point scale (1 = absence of engagement, 5 = maximum engagement) at 2-minute intervals. This approach was based on established methods for assessing flow states in creative activities (e.g., Csikszentmihalyi & Csikszentmihalyi, 1992).

2. Behavioural indicators: We collected audio-visual data to capture facial expressions, body movements, and vocal cues, which have been shown to correlate with engagement and flow states in previous research (e.g., Boccignone et al., 2018).
3. Physiological measures: We used wearable technology to measure galvanic skin response and heart rate, which can indicate arousal and cognitive engagement (e.g., Egger et al., 2019).

Such characteristics have been measured from the audio, video, inertial and physiological signals using existing methodologies (Baltrusaitis et al., 2018; Cao et al., 2019; Greco et al., 2016; Mauch & Dixon, 2014; McFee et al., 2015; Serengil & Ozpinar, 2020; Tomenotti et al., 2024). Subsequently, we synthesised these findings to construct a predictive model capable of identifying salient attributes that signal creative engagement, thus automating the detection of this experience within collaborative groups. To this purpose, we designed a feed-forward network to address group engagement as a regression problem: the multi-faceted description of the group interaction over time was fed into the neural network that returned a value, representing the predicted engagement.

3. Methodology

3.1. Participants

The study involved 12 female participants aged between 77 and 93 years (M = 84 years). All participants had no previous experience with computer devices or video calls. Participants were recruited from a local retirement home. Informed consent was obtained from all participants and the study was approved by the Institutional Ethics Commission for Research in Psychology (approval date: 11 May 2021, protocol number: 34-21).

3.2. Experimental design

We employed a within-subjects design where each participant took part in a collaborative creative task via video call. The participants were organized into groups of three, with four separate acquisition sessions conducted.

3.3. Experimental task

The collaborative task involved creating recipes using a limited set of ingredients: flour, eggs, cheese, and tomatoes. An image of these ingredients was displayed to participants during the task.

3.4. Procedure

Based on pilot testing, we determined that groups of three participants and sessions lasting approximately 20 minutes were optimal for engagement and participant comfort. Specifically, different team sizes are adopted in the literature, such as teams of 8-9 participants in Chamakiotis et al. (2013) and groups of four-five participants in Abi Saad and Agogué (2023). Initially, we formed groups of five participants plus a researcher, but this setup limited the participation. Therefore, we reduced the number of participants to three. We also planned 30-minute sessions as in Chamakiotis et al. (2013), but participants showed discomfort after 20 minutes, invalidating the experiment. Consequently, we reduced the session duration to a maximum of 20 minutes. At the start of each session, the researcher introduced the collaboration task:

> You are invited to join a team culinary challenge. Here is the scenario: guests are on their way for lunch and you find that you have not shopped for ingredients. Your cupboard contains only flour, eggs, cheese, and tomatoes. Let us brainstorm: with these items, what dishes could we create to serve our guests? Feel free to use any combination of these four ingredients to create delicious recipes. Each participant is encouraged to suggest one or more recipes, and you can also build on ideas presented by others. There is no limit to the number of recipes that you can come up with. We will take turns presenting our ideas, and then we will decide on a recipe to make together. If you have any questions or need assistance at any point, please don't hesitate to speak up.

The task was designed to be accessible and engaging for elderly participants, drawing on familiar contexts and encouraging collaboration.

3.5. Materials and apparatus

Participants used computers equipped with webcams and microphones. The Zoom platform was used for video conferencing. A mosaic video of the entire virtual meeting was recorded for synchronisation purposes. Figure 1 shows a screenshot of the online collaboration task as seen by the participants.

4. Multi-modal data: acquisition and processing

The setup for data acquisition during the virtual interaction experiences is depicted in Figure 2. In each acquisition, three participants were in their rooms at a retirement home, in front of a computer, wearing a wristband and earphones equipped with a microphone. It is worth noting that, as shown in the figure, the acquisition conditions were not tightly controlled, with variations in light, environment, and participant positions. Participants could choose their preferred seating position during the call, whether closer to or farther from the monitor. During the session, a psychologist was available in a separate room to facilitate the call.

During the sessions, we acquired multimodal signals according to three different modalities: physiological, audio, and video (see Table 1). Physiological signals were collected using the Empatica E4 bracelet (Empatica, 2019), specifically designed for research and clinical applications, and possessing medical (FDA) and electronic certifications. Participants wore the bracelet on their non-dominant hand to attenuate the effect of noise in the acquisitions. Speech recordings of all participants were obtained through video conferencing using

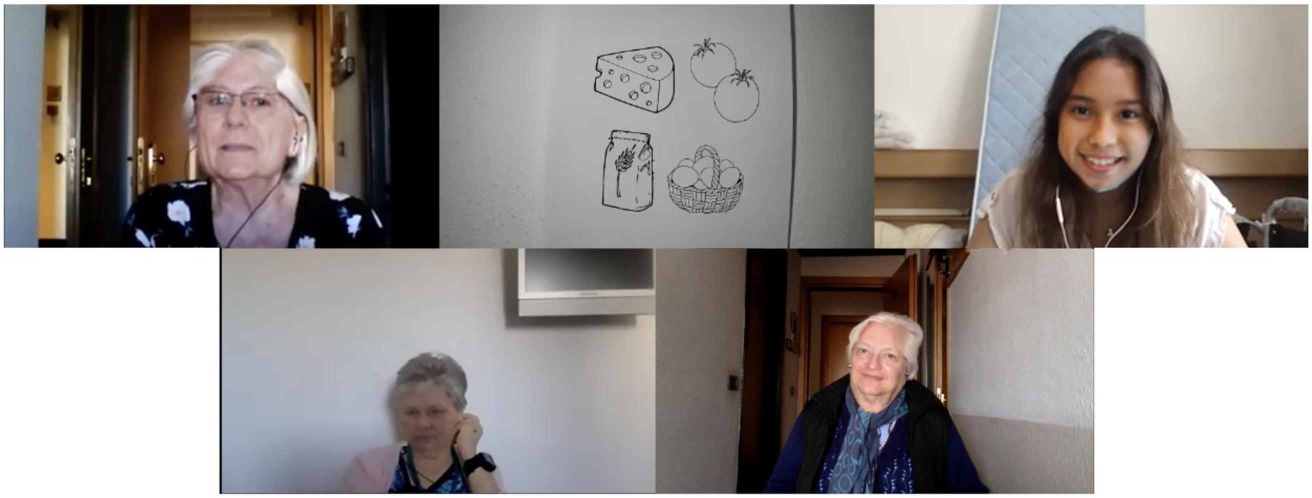

Figure 1. Screenshot example, where the identity of the participants is anonymised by exploiting DeepPrivacy model (Hukkelås et al., 2019). The Centre showcases projected images of the ingredients intended for proposing recipes; positioned on the right is the psychologist, and the remaining three individuals are the experiment participants.

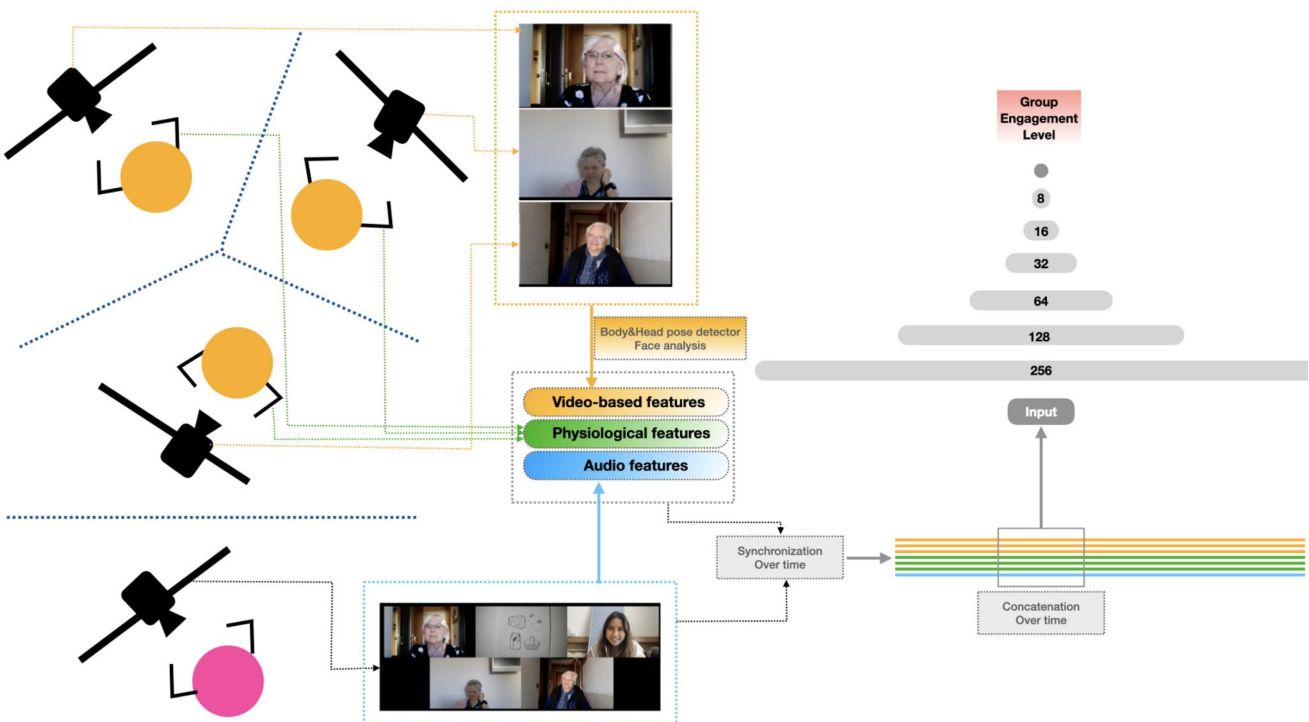

Figure 2. A visual sketch of our data acquisition. three participants (in orange on the left) participated in an online call from their rooms. They were in front of a computer equipped with a webcam and a microphone, wearing a wristband to acquire physiological measures. The acquisition was made under weakly controlled ecological conditions. In another room, the psychologist (in magenta on the bottom left) led the session. We acquired physiological, audio, and video streams, which were later synchronised and represented with an ensemble of features. These features were concatenated over time into fixed-size temporal windows and input into a neural network to predict the level of engagement.

the Zoom platform. Participants' upper-body videos were captured using their computer's webcams, while a mosaic video of the entire virtual meeting was recorded for synchronisation purposes. This approach is particularly indicated, albeit at a limited video resolution, in situations where gaining access to individual participants' computers for separate video stream recording is difficult or not feasible.

Following the acquisition and collection of the multimodal signals, synchronisation was achieved by referencing the internal clocks of the devices, and the beginning of the acquisition was automatically marked with a prearranged clapperboard signal. Next, a set of instantaneous features, as summarised in Table 1, was extracted from the synchronised signals. The following sections delve into the methodologies we employed.

4.1. Expert ratings

To assess the group's creative engagement, we opted for an approach that would capture the emergence of group flow

Table 1. List of adopted devices, recorded modalities, related sampling rate and extracted features with the relative dimensionality.

Device	Modality	Rate	Feature	Size
Webcam	Video	30 fps	Average Motion (AM)	1
	Video	30 fps	Head pose (HP)	3
	Video	30 fps	Action Units (AU)	17
	Video	30 fps	Face emotion (Em)	7
Microphone	Audio	32 kHz	Intensity (In)	1
	Audio	32 kHz	Pitch (Pi)	1
	Audio	32 kHz	Mel-Frequency Cepstral Coefficients (MC)	13
Empatica E4	Accelerometer	32 Hz	Euclidean Norm Minus One (EN)	1
	EDA	4 Hz	Skin Conductance Response (SR)	1
	EDA	4 Hz	Skin Conductance Level (SL)	1
	Heart rate	1 Hz	Heart rate (HR)	1

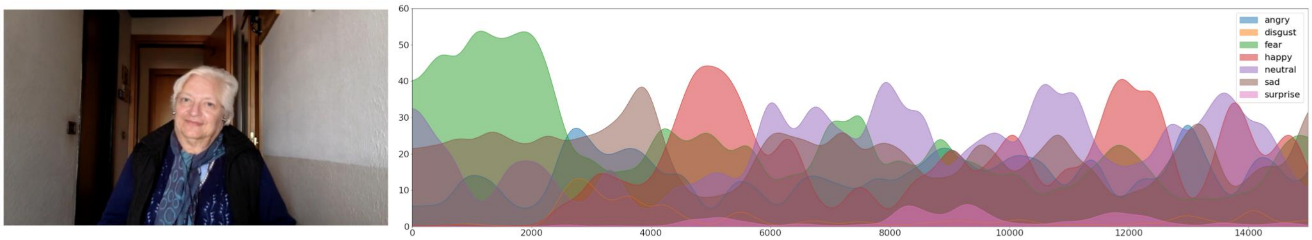

Figure 3. On the left, the experimental setting for upper-body video recording. The identity of the participant is anonymised by exploiting the DeepPrivacy network (Hukkelås et al., 2019). On the right, an example of facial emotion recognition for seven basic emotions applied over the length of the experiment for a specific participant. The activation of each basic emotion is expressed as a percentage that varies over time with respect to the overall emotional status.

behaviour through external observation rather than relying on self-reports. External raters can provide a more impartial assessment of group dynamics and engagement levels, free from individual biases or self-perception issues that might affect self-reports.

They can also observe and evaluate the entire session without interrupting the flow of the activity, which would be necessary for periodic self-reports. This decision aligns with previous research on social signal processing, such as the work by Hsiao et al. (2012), who used external raters to rate engagement levels on a 1-4 scale for arbitrary-length periods, which were then aggregated into sliding windows.

Following a similar methodology, two independent raters (each holding an M.Sc. degree in Psychology) assessed the creative engagement of each participant during each video session using a Likert-type scale ranging from 1 (absence of engagement) to 5 (maximum engagement), with assessments conducted in 2-minute intervals. To gauge inter-rater agreement, we conducted a Cronbach's alpha reliability test, revealing a high level of concordance between raters ($\alpha = 0.94$). Following this, the results were averaged to derive a group-level engagement score over time.

The overall mean engagement score across all groups was 3.61 ($SD = 1.13$) on the 5-point scale. This relatively high average score suggests that, in general, the older adult participants in our study demonstrated a good level of creative engagement in the virtual task. The standard deviation indicates moderate variability in engagement levels, which could be attributed to individual differences, group dynamics, or varying levels of comfort with technology. This finding is particularly encouraging, given that the participants had no prior experience with computer devices or video calls, suggesting that older adults can effectively engage in creative tasks in virtual environments when provided with appropriate support and a structured activity.

4.2. Behavioural indicators

4.2.1. Video features: face

The face is considered the primary communication channel that plays a crucial role in our social interactions. Therefore, we aimed to evaluate its effectiveness in this experiment by collecting facial expressiveness features, specifically the Action Units and Facial Expressions.

OpenFace (Baltrusaitis et al., 2018; Baltrušaitis et al., 2015) was used to estimate the intensity of 17 action units (specifically AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, and 45) (Zhi et al., 2020). Furthermore, we characterise facial expression using DeepFace (Serengil & Ozpinar, 2020), which categorises them into the seven basic emotions (angry, disgust, fear, happy, sad, surprise, neutral) over time, as shown in Figure 3. Although this characterisation has a limited reliability as an emotion label, it serves as a proxy for assessing the level of engagement of a subject.

4.2.2. Video features: body and motion

Considering that we were operating in an uncontrolled environment and allowing for variations in the distance between participants and the acquisition system, we also accounted for features that demanded a lower level of detail than what was typically necessary for the extraction of facial features. Furthermore, this kind of investigation could be applied under certain privacy constraints that might permit the presence of cameras while maintaining a distant viewpoint to preserve individual identities. Specifically, we considered video-based features that were more tolerant of variations in camera viewpoint, based on the analysis of body pose and motion.

To this end, all videos were processed with OpenPose (Cao et al., 2019) to obtain a representation (for each frame) of the upper body pose, detecting a set of key points corresponding to body joints or relevant locations (see a visual representation in Figure 4). For each key point, the instantaneous velocity was computed as the magnitude of the difference vector between the position of the joint in two adjacent frames. The instantaneous Average Motion (AM) was then derived as the mean of the velocity magnitude of the joints in a frame. The key points located on the face were provided as input to the HHP-net method (Cantarini et al., 2022; Tomenotti et al., 2024) to derive a representation of the head pose (HP) in terms of yaw, pitch, and roll angles expressed with respect to a reference frontal head pose (examples in Figure 4, where we show the projection on the image plane from yaw and pitch using the Tait-Bryan angles).

In Figure 5, we present examples of the evolution of the pitch angle and the average motion over time in a session, with the annotated level of engagement marked by the dotted line. The blue shadowed area represents the variability of the features among the three subjects involved in the session, while the average value is represented by the blue line. On the left, the evolution of the pitch angle shows that the level of engagement tends to be higher when the angle is close to 0.5, corresponding to the head pose of a person looking in a frontal direction. Intuitively, this indicates a high level of attention from the person. In the right plot, the average motion shows a tendency to be higher when the level of engagement is also higher, suggesting that the person tends to be more active when engaged in an interaction.

4.3. Audio features

Acoustic features can offer valuable insights into the dynamics of social interactions in an efficient and non-invasive manner (Pentland, 2004). Therefore, we decided to incorporate audio-based features into our analysis. The audio signal recorded during the entire video conference was normalised to ensure consistency among all participants, effectively reducing variations caused by different recording

Figure 4. Two examples of output obtained by applying a pose detector (OpenPose (Cao et al., 2019)) followed by head pose estimation (HHP-net, (Cantarini et al., 2022)) to the video frames. The output includes a set of key points, marking relevant body locations, and a vector indicating the rough orientation of the head. For visualisation purposes, we also show some important connections between joints, although they are not included in the output of the pose detector.

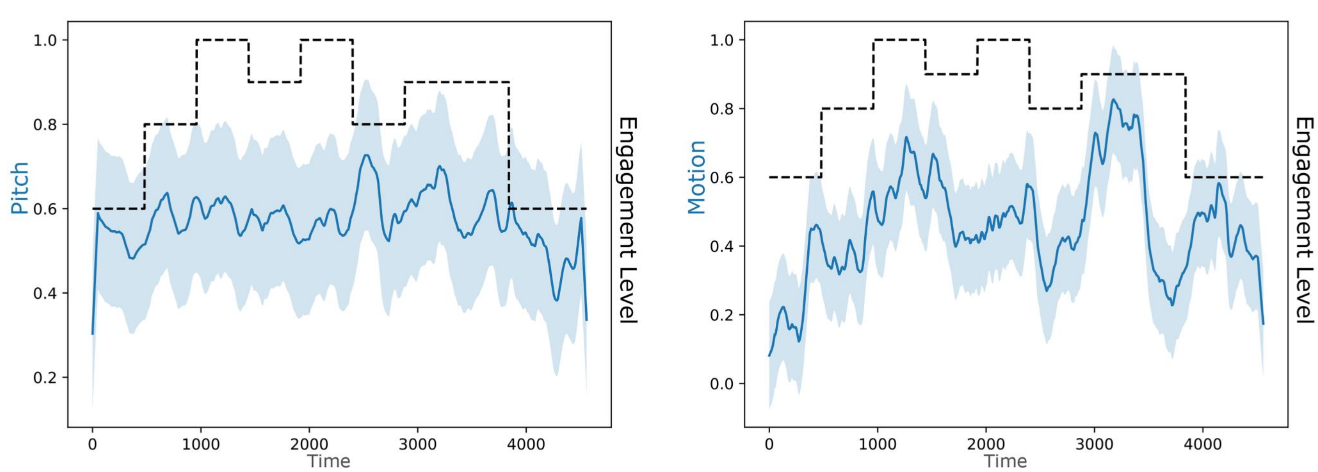

Figure 5. Examples of the behaviour of head and motion features over time. In both plots, the level of engagement is marked with the dotted line, while in blue the mean value of the features among the three subjects involved in a session is reported. The shadowed area indicates the variability between the subjects. Above, the evolution of the pitch angle is shown. The engagement is higher where the pitch is close to 0.5, corresponding to the person heading the frontal view. This is an indication of the attention of the person towards the monitor, and thus the call. At the bottom, the average motion of the body joints is reported. Here we may observe that the quantity of motion seems directly correlated with the engagement, showing that if the latter is high the person tends to be dynamically active.

environments. This normalisation process involved applying a uniform gain across the signal while preserving the signal-to-noise ratio and overall dynamic levels. We used Root Mean Square (RMS) normalisation, a type of loudness normalisation typically used for this purpose. The RMS energy signal was then extracted using the Librosa library (McFee et al., 2015). We employed a hop length of 256 samples and a frame length of 512 samples (i.e., 8 ms and 16 ms respectively @ 320 kbps), thus obtaining an energy signal at a sample rate of 125 Hz. The pitch, corresponding to the physiological parameter of the frequency of vibration of the vocal folds (fundamental frequency), was estimated using the two-step algorithm pYIN (Mauch & Dixon, 2014), followed by a low-pass filter with a cut-off at 400 Hz, given the a priori knowledge that the female fundamental frequency spans from 165 to 255 Hz (Baken & Orlikoff, 2000). Finally, the initial 13 Mel-Frequency Cepstral Coefficients (MFCC), which capture the general shape of the spectral envelope (commonly associated with timbre) through the short-term power spectrum of a sound, along with their first- and second-order derivatives, were computed.

4.4. Physiological measures

The physiological responses of an individual's body enable the detection of changes in arousal, stress levels, and cognitive workload. These factors are closely associated with the level of involvement, interest, or attention of the individual during various activities or interactions (Boccignone et al., 2018; Egger et al., 2019). Concerning each signal recorded by the Empatica E4 wristband, specific preprocessing was needed, considering the nature of each signal.

Electrodermal activity (EDA) measures electrical skin resistance in the presence of sweat produced by the body. More precisely, when significant sweating occurs, the electrical skin resistance decreases, whereas dry skin produces higher resistance. Emotions with prominent arousal, such as excitement, stress, or fear, can induce fluctuations in skin conductivity (Lang et al., 1993). An EDA signal has two main additive components: a slowly changing tonic part, referred to as the skin conductance level (SCL), and a phasic skin conductance response (SCR), characterised by rapidly changing peaks associated with short-term stimuli. To quantify the SCR amplitude, we applied a decomposition process to the EDA signal, based on the cvxEDA algorithm (Greco et al., 2016).

The number of complete heartbeats in a specific time window, referred to as heart rate (HR), is closely related to emotional arousal and depends on the activity of the sympathetic and parasympathetic nervous systems (Zhu et al., 2019). The HR signal was used as-is since it already represented a high-level characteristic of heart activity.

Another signal provided by the Empatica E4, though not strictly related to physiological cues, was from the three-axis accelerometer. This type of signal measured the participant's movement during the experiment, which could be linked to a stressful or particularly engaging state (Hooker & Masters, 2016). The raw signals from each of the three axes were combined into a vector magnitude calculated as the Euclidean Norm Minus One (ENMO), summing the squared acceleration of the three axes and then subtracting the gravitational component ($1\,g$, one gravitational unit). The increases in ENMO values were assumed to indicate higher levels of physical activity.

5. Analysis of group engagement through multimodal features

We discuss the methodology for the analysis of group engagement, with a summarised visual representation provided in Figure 2.

5.1. Features collection

To describe a virtual interaction session, we first extracted the set of features for each participant from the video and other sensors, as described in Section 4. Unlike video and physiological signals, which are distinct for each subject, the audio features were collected at a group level (see Section 4.2).

In principle, at each instant, each individual could be represented by a 47-element vector (as derived from the last column in Table 1), incorporating 28 video-based features, 4 physiological features, and 15 audio features. However, in this proof-of-concept experiment, our aim was to evaluate the reliability of different subsets of feature modalities. The identification of feature subsets was guided by envisioning various potential application scenarios, considering structural constraints (e.g., the availability of certain sensors, the placement of cameras at different distances, or difficulties in using wearable devices with uncooperative subjects) or privacy issues (e.g., the prohibition of camera use).

In general, for any feature subset (e.g., facial, audio, and physiological features alone or combined), we gathered the corresponding instantaneous feature representations over time for the subjects participating in the experimental session under analysis. Naturally, the audio features were collected only once, as they pertain to the entire group. Each feature subset aimed to capture the collective group's behaviour through the temporal measurements of the chosen features.

Before proceeding to the next step, we synchronised the signals and resampled them at a common frequency of 4 Hz (the lowest in Table 1, excluding heart rate). A sketch of this pipeline is shown in Figure 2-right.

5.2. Regression model

We addressed the task of estimating group creative engagement as a regression problem. The features of each subset, gathered and processed as explained above, were provided as input to a supervised learning algorithm. Considering the limited amount of available data, we used a Leave-One-Group-Out paradigm: given the groups available, we left out the data of one group as a test set while training the model on the data from the remaining groups. Training and test sets were created by partitioning the group features into time windows of 2 seconds (as shown in Figure 2) and reshaping them into one-dimensional vectors.

The vector was provided as input to a neural network to predict the value of group engagement at each time instant. The neural network was composed of six layers (with 256, 128, 64, 32, 16, and 8 neurons and ReLU non-linear activation) and a final one-dimensional output layer. For all experiments, the network was trained under the same conditions, using Adam as an optimiser and the square loss function, for 200 epochs with a batch size of 32. Considering that the annotations produced by the domain experts were rather sparse (being provided every 2 minutes, see Section 4.1), we opted to propagate backward the annotated engagement level to the samples between two annotation instants.

5.3. Prediction results

The results we obtained are summarised in Table 2. The investigation was carried out by evaluating different subsets of features to assess their suitability for this task. Each row in the table corresponds to a model derived from the feature subset reported in the first column. The performance of the regression model was evaluated in terms of the Mean Absolute Error (MAE) and the standard deviation between the annotations of creative engagement of the group and the predicted values (recall that engagement takes values in the interval $[1-5]$).

In the table, we present our results according to three different views. First, we provide the MAE considering the instantaneous predictions of our models. Next, we applied temporal smoothing using a simple running average over a window of approximately 15 seconds to reduce instantaneous noise in the predictions. Finally, we aggregated the predictions into 2-minute time windows that fully overlap with the annotations provided by the psychologist, thus achieving a one-to-one mapping between annotations and predictions.

Taking a broader view, we notice a consistent behaviour of different feature sets across the three columns.

Table 2. Average performance achieved by individual feature sets or their combinations.

Feature set	MAE	MAE (smoothing)	MAE (2 mins)
Face (AU, Em)	0.86 ± 0.70	0.81 ± 0.66	0.78 ± 0.27
Body (AM,HP)	1.03 ± 0.83	0.96 ± 0.83	0.98 ± 0.09
Audio (In, Pi, MC)	0.82 ± 0.65	0.80 ± 0.73	0.84 ± 0.13
Physio (SL,SR,HR)	0.94 ± 0.89	0.99 ± 0.91	0.94 ± 0.30
Face + Audio	0.84 ± 0.65	0.79 ± 0.64	0.81 ± 0.14
Face + Body	0.87 ± 0.67	$\mathbf{0.77} \pm 0.64$	0.77 ± 0.22
Body + Audio	0.83 ± 0.63	0.84 ± 0.60	0.84 ± 0.12
Physio + Face	0.93 ± 0.75	0.86 ± 0.77	0.90 ± 0.19
Physio + Body	0.95 ± 0.81	0.85 ± 0.80	0.98 ± 0.33
Physio + Audio	0.96 ± 0.74	1.02 ± 0.63	0.96 ± 0.30
Face + Body + Audio	$\mathbf{0.81} \pm 0.61$	0.78 ± 0.63	$\mathbf{0.75} \pm 0.18$
Face + Physio + Audio	0.99 ± 0.78	0.88 ± 0.75	1.00 ± 0.29
Body + Physio + Audio	0.95 ± 0.77	0.90 ± 0.76	0.97 ± 0.28
Face + Body + Physio	0.96 ± 0.77	0.91 ± 0.73	0.93 ± 0.20
Face + Body + Audio + Physio	1.01 ± 0.76	0.87 ± 0.78	1.06 ± 0.35

The outcomes are presented in the format MAE ± STD. For the column in the Centre, smoothing is applied considering time windows of length $\sim 15s$. the top-performing result per column is indicated in bold, and the second and third-best outcomes are underlined. A worthwhile reminder is that the group creative engagement values are in the range $[1, 5]$.

Furthermore, on average, the most favourable MAE results are obtained when time smoothing is applied, while a substantial reduction in standard deviation is achieved when data are aggregated every 2 minutes. In the following discussion, we mainly reference the results obtained by aggregating over two minutes (last column), as they are the most consistent with the annotations.

Upon closer examination of the distinct feature sets, it becomes evident that, when considered separately, the Face and Audio features yield the most favourable MAE results. In contrast, Body and Physiological features consistently exhibit lower performance, with an approximate 0.15 MAE deterioration. This outcome is intuitively plausible, given that facial expressions and vocal cues are the primary channels of communication that we rely on heavily during social interactions.

Furthermore, when analysing all pairs, it becomes clear that all pairs excluding the *Physio* features perform comparably, with a slightly better MAE for *Face and Body* and *Face and Audio*, once again highlighting the importance of facial features. Notably, the combination of *Body and Audio* features appears to be the most consistent across all three columns, indicating that, despite some noise, it maintains stable performance. Overall, it seems that physiological features contribute minimally to this task, a conclusion further supported when they are included in larger feature sets. Indeed, the triplet *Face, Body, and Audio* provides the best MAE among the triplets and achieves the absolute best MAE in the first and last columns.

5.4. Statistical inference

Once the comparative metrics were computed for all the subsets of features, we applied statistical significance tests to determine whether the observed differences between experiments were due to sampling error or chance. Table 2 reports the average performance achieved by individual feature sets or their combinations. The purpose of the statistical tests conducted was to establish whether these differences reflect real improvements of one feature set over another.

To achieve this, we used conventional statistical hypothesis testing procedures. For more than two populations, repeated ANOVA with Tukey's post-hoc test is typically employed. However, such a procedure requires normality and homoscedasticity (equal variance) among all populations considered in the test. When these conditions are not met, as was the case here, a Friedman test with Nemenyi post-hoc test to establish pairwise differences among populations is the appropriate approach. More specifically, the Nemenyi post-hoc test was used after rejecting Friedman's null hypothesis of equality of sample medians.

We performed these statistical tests using the *autorank* Python package (Herbold, 2020). This package analysed metrics related to instantaneous predictions (as described in Section 5.3) across all combinations of feature sets.

The results of the statistical testing procedure on the Mean Absolute Error (MAE) of the different feature sets are as follows:

1. We reject the null hypothesis that any of the populations are normally distributed.
2. We reject the null hypothesis of the Friedman test, which states that there is no difference in the central tendency of the populations.

These results indicate that there are statistically significant differences among the feature sets in terms of their predictive performance.

According to the post-hoc Nemenyi test results, we inferred the absence of significant differences within the following group comparisons: face-body-audio versus audio; audio versus body-audio; body-audio versus face-audio; face versus face-body; physio versus physio-body-audio, and physio-face-audio; physio-body-audio versus physio-face-audio, and physio-audio; physio-audio versus physio-face-body, and physio-body; and physio-face-body versus physio-body, and physio-face-body-audio. Notably, all other comparisons showed statistically significant differences.

In the Nemenyi test, we calculated the difference between the mean rankings among all the feature sets considered. Specifically, after calculating the MAE for each feature set, a ranking was established, with lower MAE values receiving higher ranks. In cases of tied MAE values, the corresponding feature sets shared the average of their ranks. The mean rank for each feature set was then computed by averaging these ranks. Consequently, a lower MAE contributed to a higher rank, and the mean ranks served as a concise measure for comparing the overall performance of different feature sets.

The output of the post-hoc Nemenyi test can be visualised through the Critical Difference (CD) diagram in Figure 6. This diagram is often used in the context of post-hoc multiple comparison tests; it visually represents the minimum significant difference required between the means of two groups to be considered statistically different. The mean ranks of each feature set were considered, where, as already mentioned, higher ranks mean lower MAE in this specific case. Each feature set is represented by a vertical line, and the horizontal bars connecting these lines indicate the critical differences. Feature sets whose difference in ranks did not exceed the significance level of $\alpha = 0.05$ are joined by thick lines and cannot be considered significantly different. On the other hand, if there is no intersection, it indicates a statistically significant difference.

By examining the equivalence groups, we can deduce that each group is defined by a distinct primary feature while incorporating various combinations of features, including the primary one. This is evident in the group with the highest mean rank, where audio emerges as the primary feature, and similar patterns can be observed in the remaining groups. For instance, physiological signals exhibit limitations in optimising the performance of any group. Furthermore, even though there is a significant difference between the groups, it can be observed from Table 1 that the mean ranks do not change much for the first six feature sets. This suggests the feasibility of implementing an experimental protocol using any of these feature sets, which results in satisfactory performance in all cases.

6. Discussion and conclusions

In this work, we contributed a methodology for the automatic assessment of creative engagement in virtual teams, with a specific focus on older adults - a demographic that has been understudied in the context of virtual collaboration.

Our primary contribution lies in advancing the empirical application of the Networked Flow model to virtual creative collaboration among older adults. By demonstrating how individual behavioural and physiological indicators can be aggregated to predict group-level creative engagement, we provide empirical support for the model's conceptualisation of collective creativity as an emergent phenomenon in this previously unexplored demographic (Gaggioli et al., 2013). This insight has the potential to bridge the gap between individual-level creativity and team-level creative outputs in virtual settings for older adults, offering a more comprehensive understanding of how creative processes unfold in digital collaborative spaces for this age group.

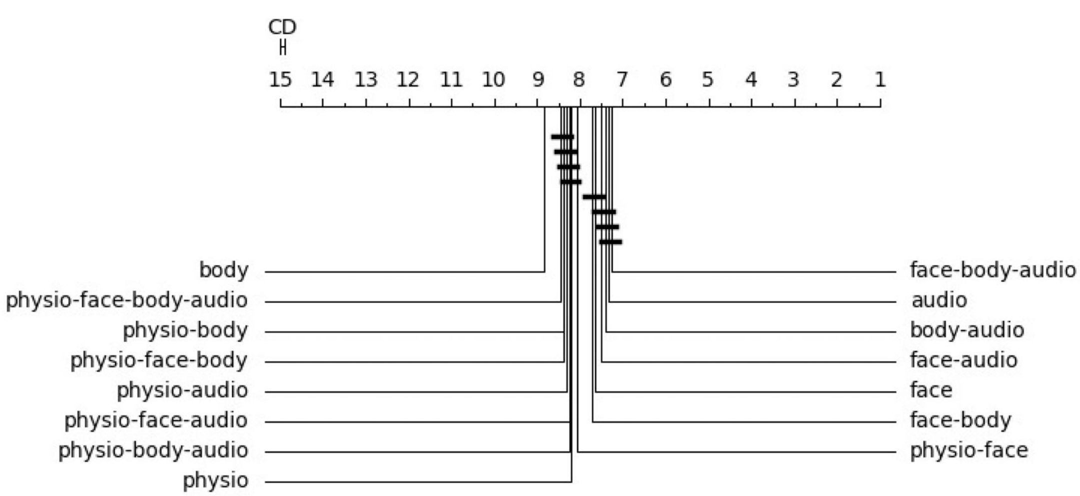

Figure 6. Result of the Nemenyi post-hoc test shown as a Critical distance (CD) diagram. CD diagrams visualise the mean ranks of populations. Populations that are not significantly different are connected by a horizontal bar.

Our second contribution is a novel multimodal approach for measuring creative engagement in virtual settings, with a particular emphasis on its application to older adults. Importantly, our study diverges from traditional methodologies by employing independent raters rather than self-reports to assess engagement levels. This approach mitigates potential biases associated with self-reporting and provides a more objective measure of engagement. In particular, our results suggested that individual creative engagement in virtual settings is manifested through observable behavioural cues, even in older adults who may be less familiar with digital technologies.

Our third contribution is demonstrating the feasibility of promoting and measuring creative engagement among older adults in digital environments - an area that has received limited attention in previous research. The overall mean engagement score across all groups indicated a good level of creative engagement among participants. Such results challenge assumptions about technology adoption and engagement among older adults, suggesting that with appropriate support and structure, this demographic can effectively participate in and benefit from virtual creative collaborations. This opens new possibilities for inclusive social activities that can be accessed regardless of physical limitations or geographical constraints, which is particularly relevant for older populations who may face mobility challenges.

Despite these contributions, our study has limitations that point to directions for future research. Our sample size was relatively small and homogeneous, consisting only of female participants from a single retirement home. Future studies should aim for larger, more diverse samples to enhance generalisability. Additionally, our experiment was conducted in a controlled setting with a specific creative task. Further research could explore how our methodology performs in more diverse contexts and with different types of creative activities.

Future research should focus on improving the precision and effectiveness of creative engagement measurement techniques for older adults, ensuring their resilience and ease of use in different contexts, and considering cultural nuances in engagement patterns. Furthermore, triangulating quantitative findings with qualitative methodologies, such as in-depth interviews with elderly participants, would provide richer insights into the cognitive and emotional aspects of using technology for social connection and its impact on mental well-being.

In conclusion, our study provides a novel approach for automatically measuring creative engagement in virtual teams of older adults, addressing an increasingly important area as our society ages and becomes more digitally connected. By combining multimodal data analysis with machine learning techniques, we describe a method that could yield reliable and easily obtainable indicators of engagement. This approach has the potential to support the development and evaluation of virtual platforms and activities tailored for older adults, ensuring that they can fully participate in and benefit from digital social and creative experiences. Moreover, these indicators could be extended to virtual rehabilitation group settings, offering valuable tools for healthcare professionals to monitor and enhance patient engagement in remote therapy sessions. Ultimately, such tools could play a crucial role in promoting cognitive health, fostering social connections, and enhancing overall well-being in later life, as virtual interactions become an integral part of elderly care, rehabilitation, and social engagement strategies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work has been supported by Fondazione Cariplo, through project "Stairway to elders: bridging space, time and emotions in their social environment for wellbeing," grant no. 2018-0858.

ORCID

Nicoletta Noceti http://orcid.org/0000-0002-6482-4768
Alice Chirico http://orcid.org/0000-0002-4587-0232
Vittorio Cuculo http://orcid.org/0000-0002-8479-9950
Giuliano Grossi http://orcid.org/0000-0001-9274-4047
Francesca Odone http://orcid.org/0000-0002-3463-2263
Andrea Gaggioli http://orcid.org/0000-0001-7818-7598
Raffaella Lanzarotti http://orcid.org/0000-0002-8534-4413

References

Aagaard, J. (2022). On the dynamics of zoom fatigue. *Convergence: The International Journal of Research into New Media Technologies*, *28*(6), 1878–1891. https://doi.org/10.1177/13548565221099711

Abi Saad, E., & Agogué, M. (2023). Creativity in virtual teams: Systematic review, synthesis and research agenda. *Creativity and Innovation Management*, *32*(1), 117–140. https://doi.org/10.1111/caim.12540

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Cengage Learning.

Baltrušaitis, T., Mahmoud, M., & Robinson, P. (2015). Cross-dataset learning and person-specific normalisation for automatic action unit detection [Paper presentation]. 2015 11th Ieee International Conference and Workshops on Automatic Face and Gesture Recognition (fg) (Vol. 6, pp. 1–6). https://doi.org/10.1109/FG.2015.7284869

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). *Openface 2.0: Facial behavior analysis toolkit* [Paper presentation]. 2018 13th Ieee International Conference on Automatic Face & Gesture Recognition (fg 2018) (pp. 59–66). https://doi.org/10.1109/WACV.2016.7477553

Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., & Lanzarotti, R. (2018). Deep construction of an affective latent space via multimodal enactment. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 865–880. https://doi.org/10.1109/TCDS.2017.2788820

Burke, M., Marlow, C., & Lento, T. (2010). Social network activity and social well-being [Paper presentation]. Proceedings of the sigchi Conference on Human Factors in Computing Systems (pp. 1909–1912). https://doi.org/10.1145/1753326.1753613

Cantarini, G., Tomenotti, F. F., Noceti, N., & Odone, F. (2022). *Hhp-net: A light heteroscedastic neural network for head pose estimation with uncertainty* [Paper presentation]. Proceedings of the Ieee/Cvf Winter Conference on Applications of Computer Vision (Wacv) (pp. 3521–3530). https://doi.org/10.1109/WACV51458.2022.00340

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., & Sheikh, Y. A. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*, 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

Chamakiotis, P., Dekoninck, E. A., & Panteli, N. (2013). Factors influencing creativity in virtual design teams: An interplay between technology, teams and individuals. *Creativity and Innovation Management*, *22*(3), 265–279. https://doi.org/10.1111/caim.12039

Chamakiotis, P., & Panteli, N. (2023). Engendering creativity in temporary virtual project teams: The case of a product design firm. In *Handbook of virtual work* (pp. 347–360). Edward Elgar Publishing. https://doi.org/10.4337/9781802200508.00028

Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (1992). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge university press. https://doi.org/10.1017/CBO9780511621956

Czikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

Egger, M., Ley, M., & Hanke, S. (2019). Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, *343*, 35–55. https://doi.org/10.1016/j.entcs.2019.04.009

Empatica (2019). *E4 wristband*. https://www.empatica.com/en-eu/research/e4/

Gaggioli, A., Mazzoni, E., Benvenuti, M., Galimberti, C., Bova, A., Brivio, E., Cipresso, P., Riva, G., & Chirico, A. (2020). Networked flow in creative collaboration: A mixed method study. *Creativity Research Journal*, *32*(1), 41–54. https://doi.org/10.1080/10400419.2020.1712160

Gaggioli, A., Mazzoni, E., Milani, L., & Riva, G. (2015). The creative link: Investigating the relationship between social network indices, creative performance and flow in blended teams. *Computers in Human Behavior*, *42*, 157–166. https://doi.org/10.1016/j.chb.2013.12.003

Gaggioli, A., Riva, G., Milani, L., Mazzoni, E., Gaggioli, A., Riva, G., … Mazzoni, E. (2013). The emergence of networked flow. *Networked Flow: Towards an Understanding of Creative Networks* (pp. 55–70). Springer. https://doi.org/10.1007/978-94-007-5552-9

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2016). cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Bio-Medical Engineering*, *63*(4), 797–804. https://doi.org/10.1109/TBME.2015.2474131

Herbold, S. (2020). Autorank: A python package for automated ranking of classifiers. *Journal of Open Source Software*, *5*(48), 2173. https://doi.org/10.21105/joss.02173

Hofer, M., & Hargittai, E. (2024). Online social engagement, depression, and anxiety among older adults. *New Media & Society*, *26*(1), 113–130. https://doi.org/10.1177/14614448211054377

Hooker, S. A., & Masters, K. S. (2016). Purpose in life is associated with physical activity measured by accelerometer. *Journal of Health Psychology*, *21*(6), 962–971. https://doi.org/10.1177/1359105314542822

Hsiao, J. C-y., Jih, W-r., & Hsu, J. Y-j (2012). *Recognizing continuous social engagement level in dyadic conversation by using turn-taking and speech emotion patterns* [Paper presentation]. Workshops at the twenty-sixth aaai conference on artificial intelligence. https://cdn.aaai.org/ocs/ws/ws0919/5341-22681-1-PB.pdf

Hukkelås, H., Mester, R., & Lindseth, F. (2019). *Deepprivacy: A generative adversarial network for face anonymization* [Paper presentation]. International symposium on visual computing (pp. 565–578). https://doi.org/10.1007/978-3-030-33720-9_44

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

Lim, C. (2023). Beyond fatiguing virtual meetings: How should virtual meetings for workplaces be supported? *International Journal of Human–Computer Interaction*, *40*(18), 1–16. https://doi.org/10.1080/10447318.2023.2231279

Lipnack, J., & Stamps, J. (2000). *Virtual teams: People working across boundaries with technology*. John Wiley & Sons, Inc.

Mauch, M., & Dixon, S. (2014). *pyin: A fundamental frequency estimator using probabilistic threshold distributions* [Paper presentation]. 2014 Ieee International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 659–663). https://doi.org/10.1109/ICASSP.2014.6853678

McFadden, S. H., & Basting, A. D. (2010). Healthy aging persons and their brains: promoting resilience through creative engagement. *Clinics in Geriatric Medicine*, *26*(1), 149–161. https://doi.org/10.1016/j.cger.2009.11.004

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and music signal analysis in python [Paper presentation]. Proceedings of the 14th Python in Science Conference (Vol. 8, pp. 18–25). https://doi.org/10.25080/Majora-7b98e3ed-003

Menary, J., Stetkiewicz, S., Nair, A., Jorasch, P., Nanda, A. K., Guichaoua, A., Rufino, M., Fischer, A. R. H., & Davies, J. A. C. (2023). Going virtual: Adapting in-person interactive focus groups to the online environment. *Emerald Open Research*, *1*(6), 1–14. https://doi.org/10.1108/EOR-06-2023-0008

Meshi, D., Cotten, S. R., & Bender, A. R. (2020). Problematic social media use and perceived social isolation in older adults: a cross-sectional study. *Gerontology*, *66*(2), 160–168. https://doi.org/10.1159/000502577

Mihaly, C. (1975). *Beyond boredom and anxiety. Experiencing flow in work and play*. Jossey-Bass Publishers.

Nadler, R. (2020). Understanding "zoom fatigue": Theorizing spatial dynamics as third skins in computer-mediated communication. *Computers and Composition*, *58*, 102613. https://doi.org/10.1016/j.compcom.2020.102613

Nguyen, M. H., Hunsaker, A., & Hargittai, E. (2022). Older adults' online social engagement and social capital: The moderating role of internet skills. *Information, Communication & Society*, *25*(7), 942–958. https://doi.org/10.1080/1369118X.2020.1804980

Ocker, R. J. (2005). Influences on creativity in asynchronous virtual teams: A qualitative analysis of experimental teams. *IEEE Transactions on Professional Communication*, *48*(1), 22–39. https://doi.org/10.1109/TPC.2004.843294

Pentland, A. (2004). *Social dynamics: Signals and behavior* [Paper presentation]. International conference on developmental learning (Vol. 5, pp. 1–5). https://vismod.media.mit.edu/pub/tech-reports/TR-579.pdf

Phillips, C. (2017). Lifestyle modulators of neuroplasticity: how physical activity, mental engagement, and diet promote cognitive health during aging. *Neural Plasticity*, *2017*(1), 3589271–3589222. https://doi.org/10.1155/2017/3589271

Serengil, S. I., & Ozpinar, A. (2020). Lightface: A hybrid deep face recognition framework [Paper presentation]. 2020 Innovations in Intelligent Systems and Applications Conference (Asyu) (pp. 23–27). https://doi.org/10.1109/ASYU50717.2020.9259802

Thomas, J. E., O'Connell, B., & Gaskin, C. J. (2013). Residents' perceptions and experiences of social interaction and participation in leisure activities in residential aged care. *Contemporary Nurse*, *45*(2), 244–254. https://doi.org/10.5172/conu.2013.45.2.244

Tomenotti, F. F., Noceti, N., & Odone, F. (2024). Head pose estimation with uncertainty and an application to dyadic interaction detection. *Computer Vision and Image Understanding*, *243*, 103999. https://doi.org/10.1016/j.cviu.2024.103999

Walsh, T. (2019). Virtual team success with the power of technology advancements. In *Advances in the technology of managing people: Contemporary issues in business* (pp. 99–107). Emerald Publishing Limited. https://doi.org/10.1108/978-1-78973-073-920191009

Zhi, R., Liu, M., & Zhang, D. (2020). A comprehensive survey on automatic facial action unit analysis. *The Visual Computer*, *36*(5), 1067–1093. https://doi.org/10.1007/s00371-019-01707-5

Zhu, J., Ji, L., & Liu, C. (2019). Heart rate variability monitoring for emotion and disorders of emotion. *Physiological Measurement*, *40*(6), 064004. https://doi.org/10.1088/1361-6579/ab1887

About the authors

Nicoletta Noceti is an Associate Professor in Computer Science at UniGE. She is a co-founder of MaLGa and PI in the Machine Learning and Vision research unit. Her research activity is mainly focused on Computer Vision and Machine Learning with a focus on methods for the understanding of human motion.

Simone Campisi earned a Master Degree in Computer Science, with a focus on AI, from the University of Genoa. He has been a research fellow at MaLGa and currently works at Imavis Srl.

Alice Chirico, PhD in Psychology, is Assistant Professor at Università Cattolica del Sacro Cuore, Milan. She leads the Experience Lab Research Group and Psychology of Music Research Unit, studying complex emotions through simulation technologies and art to explore their impact on wellbeing and cognition.

Vittorio Cuculo is an Assistant Professor at the University of Modena and Reggio Emilia, member of AImageLab research laboratory. He holds a Ph.D. in Mathematical Sciences from the University of Milan. His current research topics include human visual attention, remote physiological measurement and affective computing.

Giuliano Grossi holds a PhD in Computer Science and is currently an associate professor at the University of Milan. He was a constituent of the "Perceptual Computing and Human Sensing Laboratory" (PHuSE Lab). Current research concerns modeling and application in high performance computing for computer vision and signal processing.

Monica Michelotto is a psychotherapist at the "Istituto di Psicosomatica Integrata" in Milan and Gallarate. She specializes in treating frail elderly individuals with neurodegenerative diseases and their caregivers. She also researches Smart Healthcare and the Internet of Things, emphasizing home care usability and acceptance for the elderly and disabled.

Francesca Odone is a Full Professor in Computer Science at UniGe and a founding member of MaLGa. Her research interests include human motion analysis and scene analysis in the field of computer vision, with application to health, well-being, human-machine interaction.

Andrea Gaggioli, PhD in Psychobiology, is full professor at Università Cattolica del Sacro Cuore, leading the Research Center in Communication Psychology (PsiCom) and the International Master in User Experience Psychology (UxP). He researches digital experience psychology, exploring technology's influence on human lives through various research methods.

Raffaella Lanzarotti holds a PhD in Computer Science and is currently an associate professor at the University of Milan. She was a constituent of the "Perceptual Computing and Human Sensing Laboratory" (PHuSE Lab). Current research concerns affective computing and face analysis for clinical diagnoses.