frontiers in ICT

ORIGINAL RESEARCH
published: 24 July 2018
doi: 10.3389/fict.2018.00017

Behavioral and Physiological Responses to Visual Interest and Appraisals: Multimodal Analysis and Automatic Recognition

Mohammad Soleymani and Marcello Mortillaro*

Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Interest drives our focus of attention and plays an important role in social communication. Given its relevance for many activities (e.g., learning, entertainment) a system able to automatically detect someone's interest has several potential applications. In this paper, we analyze the physiological and behavioral patterns associated with visual interest and present a method for the automatic recognition of interest, curiosity and their most relevant appraisals, namely, coping potential, novelty and complexity. We conducted an experiment in which participants watched images and micro-videos while multimodal signals were recorded—facial expressions, galvanic skin response (GSR), and eye gaze. After watching each stimulus, participants self-reported their level of interest, curiosity, coping potential, perceived novelty, and complexity. Results showed that interest was associated with other facial Action Units than smiling when dynamics was taken into consideration, especially inner brow raiser and eye lid tightener. Longer saccades were also present when participants watched interesting stimuli. However, correlations of appraisals with specific facial Action Units and eye gaze were in general stronger than those we found for interest. We trained Random Forests regression models to detect the level of interest, curiosity, and appraisals from multimodal features. The recognition models—unimodal and multimodal—for appraisals generally outperformed those for interest, in particular for static images. In summary, our study suggests that automatic appraisal detection may be a suitable way to detect subtle emotions like interest for which prototypical expressions do not exist.

Keywords: emotion recognition, appraisal, interest, eye gaze, facial expressions, computer vision, affective computing, galvanic skin response

OPEN ACCESS

Edited by:
Nadia Bianchi-Berthouze, University College London, United Kingdom

Reviewed by:
Raymond Robert Bond, Ulster University, United Kingdom
Hongying Meng, Brunel University London, United Kingdom

****Correspondence:***
Mohammad Soleymani
soleymani@ict.usc.edu

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in ICT

Received: *30 January 2018*
Accepted: *26 June 2018*
Published: *24 July 2018*

Citation:
Soleymani M and Mortillaro M (2018) Behavioral and Physiological Responses to Visual Interest and Appraisals: Multimodal Analysis and Automatic Recognition. Front. ICT 5:17. doi: 10.3389/fict.2018.00017

1. INTRODUCTION

Interest is one of the most important, yet understudied emotions. Interest drives our focus of attention and at the same time plays an essential role for the transmission of culture (Clément and Dukes, 2013). The way we evaluate and understand our environment is deeply influenced by others' appraisals. We use information about others' interest and appreciation to build our own judgments and understand the society we are in and its cultural values. This is true for very important dimensions like religion—what is right and what is wrong—, but also for mundane activities—what is the most popular sport or the most listened song. On a smaller scale, the

detection of interest can be critical for a large number of domains, from learning to entertainment. The emotional state of a learner can influence the commitment and the outcome of the learning activity; the interest of a media consumer may predict its future preferences. So it is not surprising that automatic recognition of users' interest has broad applications. For example, a recommender system can use the level of interest as an implicit feedback to update its recommendations and improve users' satisfaction and engagement. The unobtrusive recognition of interest has applications beyond multimedia recommendation and retrieval, e.g., head-pose and gaze tracking for advertisement (Kurdyukova et al., 2012), automatic recognition of interest and boredom in education (Kapoor and Picard, 2005; Jaques et al., 2014) and video summarization (Gygli et al., 2013, 2015).

Most researchers agree that interest is an emotion and therefore is characterized by a specific cognitive appraisal structure, a consistent set of bodily expressions, and an action tendency of exploration (Izard, 2009; Mortillaro et al., 2011; Campos et al., 2013; Dukes et al., 2017). Appraisals are cognitive evaluations that occur with or without awareness - when a person faces an event or object. For example, when a person sees an image he/she evaluates the content with regard to its relevance, its novelty, its pleasantness, its compatibility with personal and social norms and whether is understandable. According to appraisal emotion theories it is the pattern of appraisal outcomes that determines the emotion that is experienced by the person (Scherer, 2009). For example, fear is typically elicited by a relatively unexpected, relevant, and negative event for which the person has a low sense of coping (Scherer, 2005). Silvia identified the appraisals of coping potential and novelty-complexity to be the typical appraisals that define the experience of interest (Silvia, 2005). People experience interest toward objects or events that are novel and at the right level of complexity - that is, not too easy or too complex to understand. Generally people who are more familiar with the subject have a higher level of interest in more complex forms of the stimuli, probably because the higher expertise requires a higher level of complexity for an optimal sense of coping (Silvia et al., 2009). In terms of appraisals, it is also important to note that interest is not always about positive objects or events, and unpleasant experiences might elicit interest (Silvia, 2008). Even if in some cases the object of interest may be intrinsically unpleasant, most authors consider interest a positive emotion: the experience of interest is subjectively pleasant or at least accompanied by an approach tendency toward the object (Ellsworth and Smith, 1988; Mortillaro et al., 2011; Campos et al., 2013).

1.1. Expressions of Interest

Research on the expression of interest has largely focused on the face. In its pioneering works Izard (2009) provided what is probably the only description of a possible prototypical facial expression of interest "Brows can either be raised, in normal shape, or drawn together and/or slightly lowered; eyes may be either widened, roundish in appearance, or squinted; cheeks may be raised; mouth can be opened and relaxed, tongue may be extended beyond the gum line, lips may be pursed" (Matias et al., 1989). More recently, Reeve and colleagues investigated which facial movements were consistently correlated with self-report of interest while participants watched films or played spatial-relations puzzles (Reeve, 1993; Reeve and Nix, 1997). Even though some behaviors seem correlated to interest—such as eye closure and eyeball exposure—they concluded that there is not a stable cluster of facial signs to express interest. Conversely, in a recent study, Campos et al. (2013) listed two Facial Action Units (AU) that were strongly associated with the expression of interest (inner brow raise and brow lowerer), plus five facial and head movements that were weakly associated. All in all, the evidence about the existence of a prototypical facial expression of interest seems weak. This conclusion however does not imply that there is no movement in the face or in the body that is consistently related to the experience and the communication of interest. Mortillaro et al. (2011) suggested that the specific movements able to differentiate among positive emotions should be looked for in the unfolding of the expression, in its dynamics, and not in a static facial configuration. Dukes et al. (2017) explored this possibility and found that indeed adding dynamic information significantly increased the accuracy with which the facial expressions of interest were recognized by others, and this effect was larger for interest than for other emotions. Importantly, when viewers could see only a static facial expression, the accuracy was only 29% (a figure that should be compared to an accuracy of 68% on average for the other emotions). This confirms that the expression of interest cannot be grasped by one static photo but may require dynamic information. Researchers investigated other expressive modalities beyond facial expression, in particular the bodily movements that accompany the experience of interest. The body may indeed play a special role in communicating specific positive emotions (Tracy and Robins, 2004; Gonzaga et al., 2006; Dukes et al., 2017; Mortillaro and Dukes, 2018). A survey on body expression and perception identified body postures and head pose to be related to expression of interest and boredom (Kleinsmith and Bianchi-Berthouze, 2013). In the study by Campos and colleagues reported before (Campos et al., 2013), interest was associated to two head movements (head tilt and head forward) and one postural shift (forward lean). Other studies have confirmed that the expression of interest—and the reaction to a novel stimulus—includes an orienting response and an approach tendency toward the object of interest. For example, in babies and children a novel stimulus is likely to produce a freezing response: babies stop all other movements and orient their attention toward the stimulus (Scherer et al., 2004; Camras et al., 2006). Kleinsmith and Bianchi-Berthouze (2013) found that body posture and head pose can be used to recognize concentration in subjects watching video game replays. In their work concentration includes interest and focus of attention.

More recently, Dael et al. (2012) investigated the bodily expressions of 12 emotions in one of the largest studies on this subject available to date. Regarding interest, they found a very distinctive pattern that could be related to the orienting response that is part of the experience of interest. More precisely, the pattern includes arms resting at side, trunk leaning forward, and asymmetrical one-arm actions. Dukes et al. (2017) compared the

nonverbal expression of interest in the body and the accuracy with which the emotion was recognized with that of five basic emotions. The results showed that, when the observer could see the body in addition to the facial expression, interest was as well recognized as the other emotions. It may therefore be that interest, like pride, may require more information than facial movements to be effectively expressed and recognized (for example head tilt and forward leaning posture Campos et al., 2013). In particular, one can speculate that the expression of interest starts with a freezing response, followed by an approach movement.

1.2. Detection

Out of the hundreds of studies and papers describing the implementation of a computer-based system for the automatic detection of emotions, only a handful of attempts included interest as one of the target emotions. Kurdyukova et al. (2012) setup a display that could detect the interest of the passersby by detecting their facial expressions and head pose. Gatica-Perez et al. (2005) proposed a system to recognize the level of interest in a group meeting from audiovisual data. A dataset of audiovisual recordings from scripted or posed meetings was annotated for the moments of interest, e.g., the moments that people were attentive and took notes. The audio channel was the most informative modality in their setting and dataset.

Detection of interest and knowledge-related emotions such as curiosity and boredom has been studied in the context of online tutoring and education. Kapoor et al. (Kapoor et al., 2004; Kapoor and Picard, 2005) used game state, body posture, facial expressions and head pose to detect interest in children playing an educational game. Body posture was sensed by a grid of pressure sensors installed on the chair where the child was sitting. They could accurately detect interesting situations during the game play with a recognition rate of 86%. Body posture was the most informative modality for interest detection. Posture, facial expressions and speech was used to recognize boredom and curiosity in learning scenarios (D'Mello et al., 2007). Sabourin et al. (2011) used a dynamic Bayesian network and multiple users' explicit input and interaction logs to recognize curiosity and boredom. Jaques et al. (2014) used eye gaze patterns to recognize curiosity and boredom in a learning scenario. Bixler and D'Mello (2013) used key strokes, pause behavior, task duration and task appraisal for detecting engagement and boredom in an essay writing task.

The most comprehensive study on recognition of interest was done by Schuller et al. (2009) who recorded a corpus of audiovisual spontaneous expressions of interest (AVIC). In their study, the experimenter and the participant were sitting on opposite sides of a table. The experimenter played the role of a marketer presenting a product to the participant. The participant was encouraged to engage in a conversation and ask questions. Audiovisual data were recorded and the segmented speaker and subspeaker turns were annotated by the degree of interest on a five points scale. The five degrees of interest were from disinterest to curiosity. Speech and non-linguistic vocalizations were transcribed and labeled by human transcribers. Across different modalities, acoustic features were shown to perform the best. Despite these few notable exceptions, the existing work on automatic emotion recognition from facial and body responses mainly focus on the recognition of few emotions with prototypical expressions (Calvo and D'Mello, 2010; Weninger et al., 2015). Recognition of emotions using continuous dimensions, such as valence and arousal, has been also explored but these attempts did not include interest (Hatice et al., 2011; Sariyanidi et al., 2015). Mortillaro et al. (2012) proposed that emotion recognition should be done through recognizing cognitive appraisals. If we recognize the appraisals as constructing factors of emotions, we can move beyond the current methods which are mainly based on the automatic recognition of prototypical expressions. Further support to this approach comes from the recent work of De Melo et al. (2014) who conducted a series of experiments in which demonstrated that appraisals are recognized by viewers and mediate the effects of emotion displays on expectations about others' intentions. In their discussion of the implications of these results, the authors argue that appraisal-based-approaches could be very useful to design human-computer-interaction systems (see also Scherer et al., 2018).

In this work, we want to first look into how the behavioral patterns of interest are different from other positive emotions. We then aim at analyzing the physiological and behavioral responses of interest, curiosity and appraisals associated with interest, namely, coping potential and novelty-complexity in the context of watching visual content (Silvia, 2006). We attempted automatic detection of interest and its related appraisals through behavioral and physiological responses. For this we used a multi-method approach in which we combined self-reports, visual recordings of participants' expressions while they were watching and looking at the stimuli, eye gaze tracking, and physiological measures. We trained an ensemble regression model, i.e., Random Forests, for detecting appraisals, curiosity and interest from facial expressions, eye gaze and GSR. The modalities were fused at decision level to create a multimodal model for detecting appraisals and interest. To the best of our knowledge, we are among the firsts to report on automatic recognition of appraisals.

2. DATA COLLECTION

2.1. Stimuli Content

In a preliminary study, a diverse set of 1005 Creative Commons licensed pictures were selected from Flickr[1] covering various topics and emotional content including people, scenery, erotic pictures, animals and celebrities (Soleymani, 2015). We paid special attention in collecting pictures diverse in their content, aesthetics and quality. We labeled the pictures through crowdsourcing on Amazon Mechanical Turk (MTurk)[2]. Each image was rated by 20 participants on interestingness, comprehensibility (coping potential), pleasantness, aesthetics arousal, complexity and novelty. Eighty images were selected as stimuli for the current work to cover the whole spectrum in terms

[1] http://www.flickr.com
[2] http://www.mturk.com

of average interestingness, pleasantness and coping potential. Images were resized from their largest version available to 1,440 × 1,080 pixels. Examples of the stimuli are given in **Figure 1**.

One hundred and thirty-two micro-videos in GIF format from Video2GIF dataset (Gygli et al., 2016) were randomly selected and annotated on similar scales on MTurk. In our experiments, we displayed the images in full screen mode. GIFs do not have adequate resolution when displayed in a full-screen mode. Hence, we extracted the higher quality equivalent from the source YouTube videos and re-encoded them to our desired format (1,920 × 1,080) with no sound. Forty micro-videos were selected to cover the whole spectrum in terms of average interestingness, pleasantness and coping potential. We opted for using GIFs due to their short duration, unimodality (only visual) and higher level of engagement (Bakhshi et al., 2016). GIFs are not always encoded with the same frame rate as the original video and contain loopiness, therefore we re-encoded them in 1.5x speed and repeated the sequence twice. Micro-videos were in average 11 s long.

2.2. Apparatus and Protocol

The experiment has received ethical approval from the ethical review board of the Faculty of Psychology and Educational Sciences, University of Geneva. Fifty-two healthy participants with normal or corrected to normal vision were recruited through campus wide posters and Facebook. From these 52 participants, 19 were male and 33 were female. Participants were in average 25.7 years old (*standard deviation* = 5.3). Participants were informed about their rights and the nature of the experiment. They then signed an informed consent form before the recordings. They received CHF40 for their participation.

Experiments were conducted in an acoustically isolated experimental booth with controlled lighting. Video was recorded using an Allied Vision[3] Stingray camera at 60.03 frames/second with 780 × 580 resolution. Stimuli were presented on a 23 inches screen (1,920 × 1,080) and participants were seated approximately 60 cm from the screen. Two Litepanels[4] daylight spot LED projectors were used for lighting participants' faces to reduce possible shadows. An infra-red block filter was mounted on the lens to remove the reflection of the infra-red light from the eye gaze tracker. Video was recorded by Norpix Streampix software[5]. Eye gaze, pupil diameter and head distance was recorded using a Tobii[6] TX300 eye gaze tracker at 300 Hz. GSR was recorded using a Biopac[7] MP-36 at 125Hz through electrodes attached on distal phalanges of index and middle fingers. Experimental protocol was run by Tobii Studio and the recordings were synchronized by a sound trigger that marked the frames before each stimulus for the camera. The same trigger was converted to a TTL trigger using a Brain Products StimTrak[8] and recorded alongside the GSR signals. To simplify the interface, we only provided the participants with a keyboard with numerical buttons that they could use to give ratings (1–7). A picture of the experimental setup is shown in **Figure 3**. Examples of facial expressions in extreme conditions of interest and disinterest are given in **Figure 2**.

Participants were first familiarized with the protocol and ratings, in a dummy run. Participants looked at each image for five seconds and then used seven point semantic differential scales to rate their interestingness (from uninteresting to interesting), invoked curiosity (how much they like to watch or look at similar content), perceived coping potential (average of two scales, easy to understand-hard to understand and incomprehensible-comprehensible), novelty (from not novel to novel), and complexity (from simple to complex). In the second part of the study, they watched the 40 micro-videos and answered the same self-report questions.

2.3. Self-Reports

Rater consistency was calculated to check the reliability of the scales. The calculated Chronbach's alpha shows strong reliability for all self-reported scores. The correlation coefficient between different ratings of 80 images and their reliability scores are given in **Table 1**. As expected interest and curiosity have a very high correlation. Coping potential and complexity are also highly correlated which means participants found the more complex stimuli less comprehensible. Despite the findings of Silvia (2006) that coping potential is a critical appraisal for interest, in our current data coping potential is not positively correlated with interest. Rater consistency scores and between-rating correlation coefficients followed a very similar pattern for ratings given to micro-videos.

The histogram of the self-reported scores to images are given in **Figure 4**. We succeeded in eliciting a wide range of interest, curiosity and novelty. However, the distribution of coping potential and complexity scores are more uneven. This is probably due to the nature of our stimuli (images) that in most cases are comprehensible and not complex. Responses to micro-videos follows a very similar distribution.

The experimental sessions were rather short (about 25 min) however the tasks were repetitive. To control for potential effects of boredom or fatigue, we calculated the correlation between interest scores and the order by which the content was displayed to the subjects. We did not find any significant correlation between them which demonstrated that fatigue did not have a systematic effect on the self-reported interest in visual content.

The database recorded—except videos from subjects' faces—are available for academic research at [http://cvml.unige.ch/resources]. For face videos, the landmarks, features and Action Units extracted will be provided for the benefit of the community.

[3]https://www.alliedvision.com/
[4]http://www.litepanels.com
[5]https://www.norpix.com
[6]http://www.tobii.com/
[7]https://www.biopac.com/
[8]http://www.brainproducts.com/

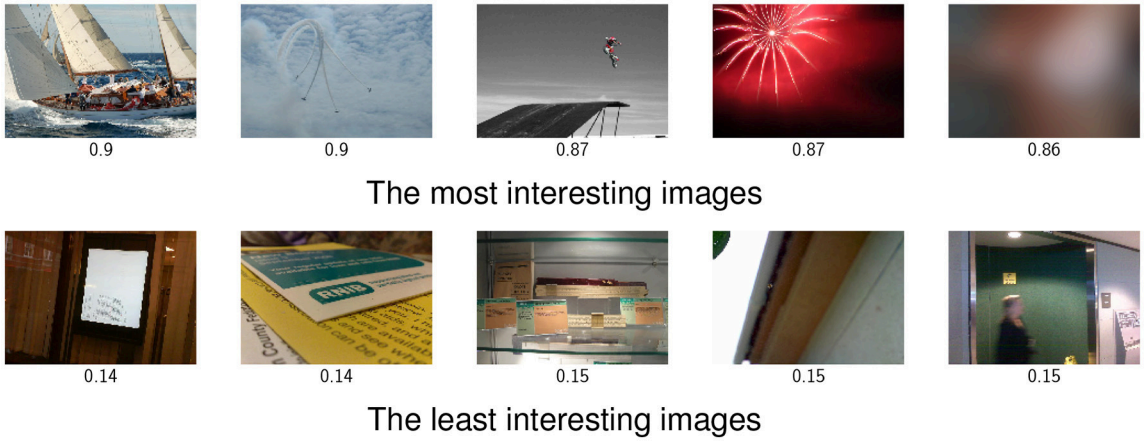

FIGURE 1 | The most and least interesting images from the pilot study. The average normalized ([0, 1]) interestingness scores are written underneath. The top right image contains explicit content and is blurred.

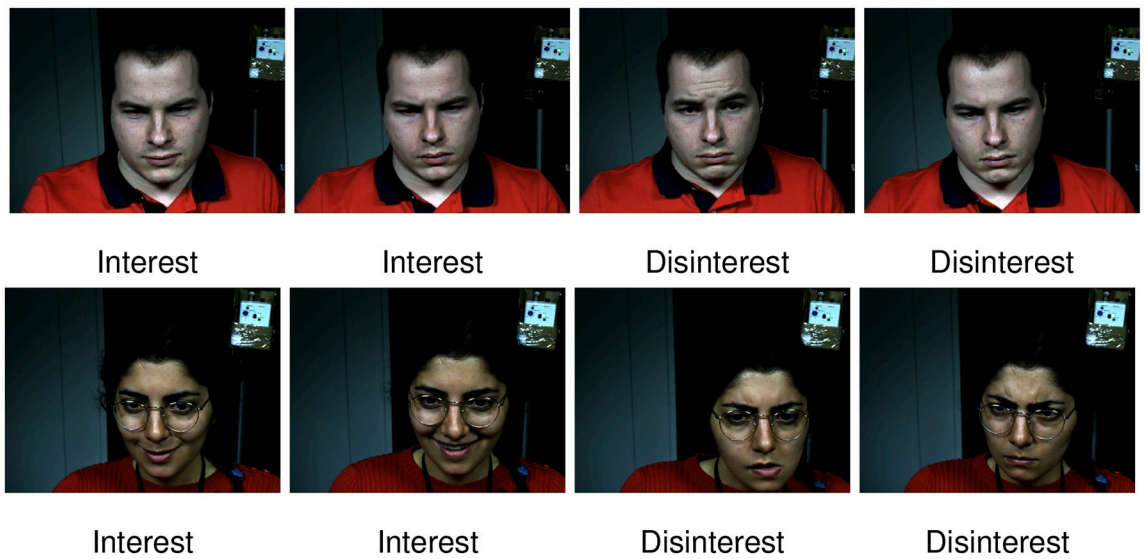

FIGURE 2 | Examples of expressions for interest (*score* = 7) and disinterest (*score* = 1). Participants facial expressions are not consistent. Expressions of the participant in the second row resembles valence more than interest; i.e., for this participant, pleasantness was the main factor for interestingness. Participants provided their written informed consent for publishing their images.

3. EXPRESSIONS AND BODILY RESPONSES OF INTEREST AND APPRAISALS

3.1. Facial Expressions

The data from two participants had to be discarded due to the technical failure in recording and synchronization. Head pose, head scale and eye gaze coordinates were extracted in addition to the facial Action Units (Ekman and Friesen, 1978). The intensity of the following Action Units were detected at frame level by OpenFace (Baltrusaitis et al., 2015, 2016): AU1 (Inner eye brow raiser), AU2 (Outer eye brow raiser), AU4 (brow lowerer), AU5 (Upper lid raiser), AU6 (Cheek raiser), AU7 (Eye lid tightener), AU9 (Nose Wrinkler), AU10 (upper lip raiser), AU12 (Lip corner puller), AU14, AU15 (Lip corner depressor), AU17 (Chin raiser), AU20 (Lip stretcher), AU23 (Lip tightener), AU25 (Lips part), AU26 (Jaw drop) and AU45 (Blink). OpenFace tracks 68 landmarks on the face (see **Figure 5**). After rotating the two-dimensional landmarks from faces to a frontal position and discarding their third dimension, we registered them to a standard face via a rigid transformation calculated by Procrustes analysis on shapes from each frame. We extracted 47 dynamic points on eyes, lips and eyebrows and used their coordinates as features for each frame. The following seven functionals were applied to the features in each trial for pooling: mean, standard deviation, median, maximum, minimum, first and third

FIGURE 3 | The recording setup including an eye gaze tracker, front-facing camera capturing face videos and galvanic skin response. Participant in this figure provided her written informed consent for publishing her image.

TABLE 1 | Chronbach's alpha rater consistency scores and Spearman rank correlation coefficients between the ratings.

Scale	Interest	Coping	Curiosity	Novelty	Complexity
Interest	–	−0.11	0.75	0.31	0.27
Coping	–	–	−0.04	−0.44	−0.69
Curiosity	–	–	–	0.26	0.20
Novelty	–	–	–	–	0.47
Cronbach's α	0.91	0.97	0.92	0.95	0.96

quartiles. This resulted in a feature vector with 658 elements for each trial. We opted for using landmarks as features since automatic Action Unit detection has a lower accuracy.

We calculated the Pearson rank correlation between the Action Units (averaged over each trial) and the ratings. In general correlations were low. The three most highly correlated Action Units with each scale are given in **Table 2**. Interest has the highest correlation with Action Units associated with positive emotions (enjoyment smile, AU6 + AU12) and negatively with AU4 (frowning), frequent in negative emotions. Curiosity has a similar pattern. Our results confirm other recent studies that found interest correlated to smiling behavior and in general being part of the category of positive emotions (Mortillaro et al., 2011; Campos et al., 2013; Dukes et al., 2017). An alternative explanation would be that we did not have extremely unpleasant stimuli that could have elicited interest but not a smiling behavior.

Coping potential and complexity are correlated with AU5 which is the eye lid raiser. The more complex or challenging the stimulus, the wider the eyes became: Eyeball exposure may be linked to the search for more visual information and previous studies suggested that it may be related to the experience of interest (Reeve, 1993; Reeve and Nix, 1997). Novelty is surprisingly associated with AU14 (dimpler) and AU23 lip tightener in addition to AU5. The presence of AU14 might be just due to chance or error in AU detection. It is also worth noting that the software we used does not have a high accuracy in detecting all Action Units and these results are not comparable with the studies in psychology with manual Action Unit coding (Mortillaro et al., 2011; Campos et al., 2013; Dukes et al., 2017).

Recent work on the expression of interest clearly found that the dynamics of expressions are important in recognition of interest (Dukes et al., 2017). Therefore, we looked at how the facial expressions unfolded during the image watching trials with the highest (>5, rated on seven-point sale) and lowest self-reported interest (<3) (see **Figure 6**). The samples with higher interest include higher activation of AU6 and AU12 (lower half of **Figure 6**), clearly indicating that interesting stimuli were associated with smiling behavior. Unlike Campos (Campos et al., 2013) we do not observe higher activation of AU1 (inner brow raiser) throughout the tasks. AU1 was more activated in interesting stimuli compared to non interesting stimuli only for the first second of the image viewing tasks; we can interpret this as the first reaction to a novel interesting stimulus, a reaction that, as the stimulus become "known" tend to disappear. Importantly, AU7 (lid tightener) is on average more activated in response to highly interesting stimuli than low interesting ones throughout the expression, confirming previous finding by Mortillaro and colleagues, who found AU7 present in 90% of the expressions of interest that they analyzed (Mortillaro et al., 2011). AU7 is very similar to eye closure, which was suggested as a marker of interest by Reeve (1993) and could be related to the focusing of the attention and the cognitive effort entailed by the experience of interest (Silvia, 2008).

3.2. Eye Gaze and Posture

Optical eye gaze trackers track the direction of gaze and provide the projected gaze. Eye gaze features such as fixations and saccades were extracted by the eye gaze analysis software, Tobii Studio. Fixations are the points where eye gaze is maintained for a minimum amount of time of 100 ms. Saccades are the eye movements between fixations. The absolute direction of saccades (measured by their absolute angle) and the relative direction with regard to the last saccade were calculated by Tobii studio. With a simplifying assumption of straight saccadic movements, we defined the scan path as the direct path between the consecutive fixations. In eye gaze analysis, often times an area of interest (AOI) is defined to study the gaze pattern locally. We defined AOI as the exact coordinates of the photos displayed; there was no AOI for the videos because they were displayed in full screen. The eye gaze tracker also record head distance from the screen. We used head distance as a measure of body posture, because it indicates whether the person is leaning forward or backward with respect to his standard resting position. Inspired by the relevant literature on interest, boredom and emotions (D'Mello

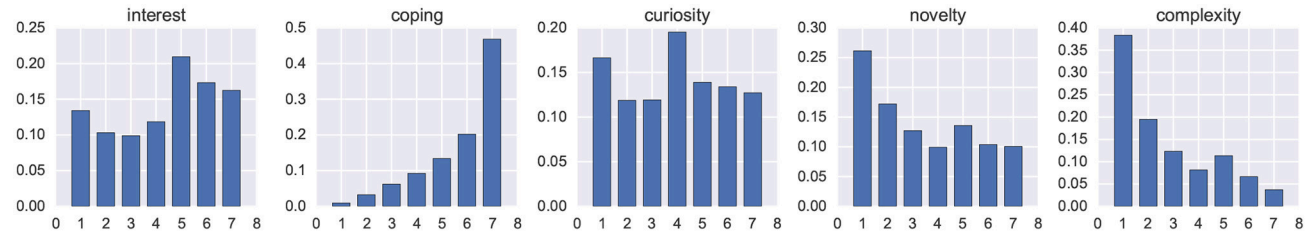

FIGURE 4 | The normalized histograms of the self-reported scores in response to images.

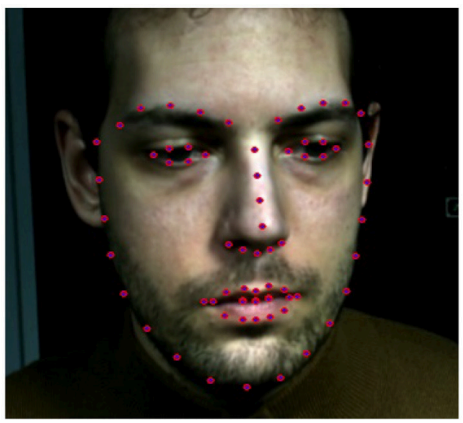

FIGURE 5 | An example of cropped detected face and its tracked facial landmarks overlaid on the original image. Participant in this figure provided his written informed consent for publishing his image.

TABLE 2 | Top three most correlated Action Units (AU) with five scales.

Scale	AU	ρ	AU	ρ	AU	ρ
IMAGES						
Interest	AU4	−0.07	AU6	0.07	AU10	0.07
Curiosity	AU12	0.07	AU6	0.07	AU23	0.07
Coping	AU5	0.13	AU14	−0.10	AU6	−0.08
Novelty	AU5	0.13	AU23	0.13	AU17	0.08
Complexity	AU5	0.15	AU4	−0.09	AU1	−0.08
MICRO-VIDEOS						
Interest	AU12	0.11	AU2	0.11	AU4	−0.11
Curiosity	AU12	0.14	AU6	0.12	AU15	0.11
Coping	AU5	0.10	AU14	−0.10	–	–
Novelty	AU23	0.16	AU6	0.15	AU14	0.14
Complexity	AU5	0.12	AU4	−0.12	–	–

ρ: Spearman rank correlation coefficient. Only correlation coefficients whose absolute value is superior to 0.05 are included ($p < 0.0001$).

et al., 2012; Soleymani et al., 2012; Blanchard et al., 2014), 60 features were extracted (see **Table 3**).

We calculated the correlation of interest, curiosity, coping, novelty, and complexity with the features extracted from the eye gaze behavior (see **Table 4**). Novelty and curiosity are associated with longer saccades; interesting and novel stimuli call for more visual exploration, a result that is in line with our finding from facial analysis, where the eye lid raiser Action Unit was correlated with the appraisal of novelty.

We compared how the head (or gaze) distance changed over time in the most interesting (>5 on seven point scale) and the least interesting images (<3). We found that the most interesting images were characterized by a shorter distance between the eyes and the screen, implying a lean forward head and/or upper body posture (see **Figure 7**). Head distance can be changed both by learning forward and head pose variations, but both these movements are indicative of an approach tendency; these results are perfectly in line with previous findings by Campos (Campos et al., 2013) who found that interest was associated with two head movements (head tilt and head forward) and one postural shift (forward lean). Importantly there is no difference in gaze distance at the beginning of the trial between high interesting and low interesting stimuli, only after 1 second the difference appears, when the participants move backwards, disengaging from low interesting stimuli. The dynamic behavior observed in the interesting stimuli seems very similar to the recent observation of Dukes et al. (2017) who suggested that the bodily expression of interest consists of two subsequent movements, freeze and then approach.

3.3. Galvanic Skin Response (GSR)

Lang et al. (1993) found that interest in images was strongly correlated with arousal. GSR is a measurement of electrical conductance on skin through a pair of electrodes and it is extensively used in psychology and affective computing to estimate someone's level of physical activation (or arousal) (Jennifer and Rosalind, 2000; Kim and André, 2008; Calvo and D'Mello, 2010; Kreibig, 2010). Indeed, skin's electrical conductance measured by GSR fluctuates with the activity of sweat glands which are driven by the sympathetic nervous system. GSR responses consists of tonic (slow) and phasic (fast and often event-related) responses. Importantly, GSR cannot be directly used to distinguish between different emotions, e.g., elation vs. disgust, but provides a measure for detecting the presence and intensity of emotions. We used the open source TEAP toolbox[9] (Soleymani et al., 2017) to extract nine features from the GSR signals. In order to capture the phasic responses, we extracted the peaks that appears in GSR signals and calculated

[9]https://github.com/Gijom/TEAP

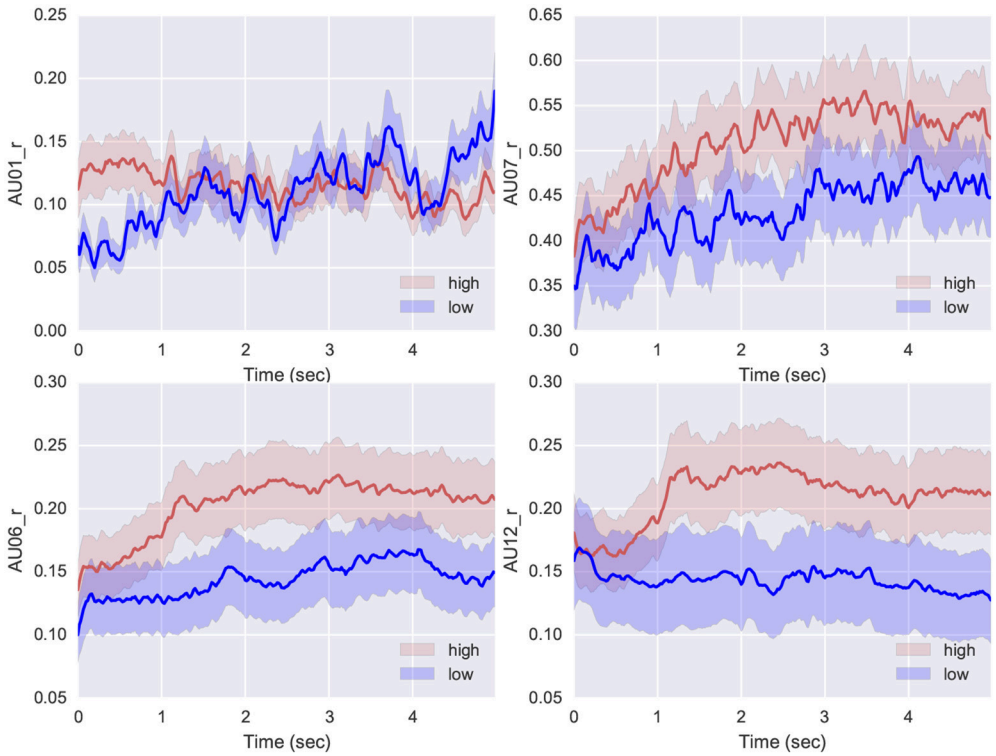

FIGURE 6 | The facial Action Units detected over time in high and low interest responses to images. The shadows represent the 95% confidence intervals. AU1, inner brow raiser; AU6, chick raiser; AU7, lid tightener; and AU12, Lip Corner Puller.

TABLE 3 | The list of 60 eye gaze features.

Feature type	Description	#
AOI	Number of fixations in the AOI, the proportion of gaze duration in AOI	2
Fixation	Number of fixations, statistical descriptives on fixation duration	8
Saccade	Number of saccedes, statistical descriptives of saccade duration, absolute and relative saccadic directions	22
Scan path	Statistical descriptives of scan path distances and their speed	14

The functionals or statistical descriptives are mean, standard deviation, first and third quartiles, median, maximum and minimum. AOI, area of interest.

their frequency of occurrence, amplitude, and rise time. Statistical descriptives were also extracted that captures both tonic and phasic characteristics of electrodermal responses. The list of features are given in **Table 5**.

We calculated the correlation between GSR features and participants' self report measures. We only found significant correlations between interest and curiosity and GSR features. Mean GSR was inversely correlated with interest ($\rho = -0.09$, $p < 0.0001$) while watching images, and number of peaks were correlated with interest while watching micro-videos ($\rho = 0.09$, $p < 0.0001$). It is also worth noting that due the limited length of the GSR signals (only five seconds for images and ~11 s for micro-videos), we could not record the possibly slower electrodermal responses that occurred outside this time window.

4. APPRAISAL AND INTEREST DETECTION

In this section, we report the results of appraisal and interest detection on responses to images and micro-videos separately. We used an ensemble regression model, Random Forests, with 100 trees and minimum leaf size of five for detecting the level of interest, curiosity and appraisals. The strength of such an ensemble method is its lower susceptibility to over-fitting. In our preliminary experiments, Random Forests outperformed Support Vector Regression with a Radial Basis Function kernel. Due to the ordinal nature of the scores, we opted for rank-normalization for labels from each participant. In rank-normalization, all the values are sorted and then the rankings are converted to values between zero and one. Features were normalized by subtracting their mean and dividing by their standard deviation. In the same manner, GSR signals were first normalized per-person to alleviate the between-participant differences. We used a 20-folding inter-participant (not participant-independent) cross-validation strategy for evaluating the regression results on five different scales, namely, interest, curiosity, coping potential, novelty

TABLE 4 | Top three most correlated eye gaze features and five scales.

Scale	Feature	ρ	Feature	ρ	Feature	ρ
IMAGES						
Interest	–	–	–	–	–	–
Curiosity	–	–	–	–	–	–
Coping	Fixation duration 1st quart.	0.07	Fixation duration median	0.07	–	–
Novelty	Gaze distance stdev.	0.116	–	–	–	–
Complexity	Fixation duration 1st quart.	0.106	Fixation duration median	0.103	Fixation duration 3rd quart.	0.088
MICRO-VIDEOS						
Interest	Pupil diameter min.	−0.153	Pupil diameter 1st quart.	−0.141	Pupil diameter mean	−0.133
Curiosity	Saccade duration 3rd quart.	−0.126	Pupil diameter min.	−0.120	Pupil diameter 1st quart.	−0.116
Coping	Saccade duration 3rd quart.	0.090	–	–	–	–
Novelty	Saccade duration max.	0.166	Saccade duration stdev.	0.139	Saccade relative direction min.	−0.089
Complexity	pupil diameter stdev.	0.105	Saccade duration max.	0.103	Saccade duration stdev.	0.087

ρ: Spearman rank correlation coefficient. Only correlation coefficients whose absolute value is superior to 0.05 are included ($p < 0.0001$). stdev, standard deviation; quart., quartile; max., maximum; min., minimum.

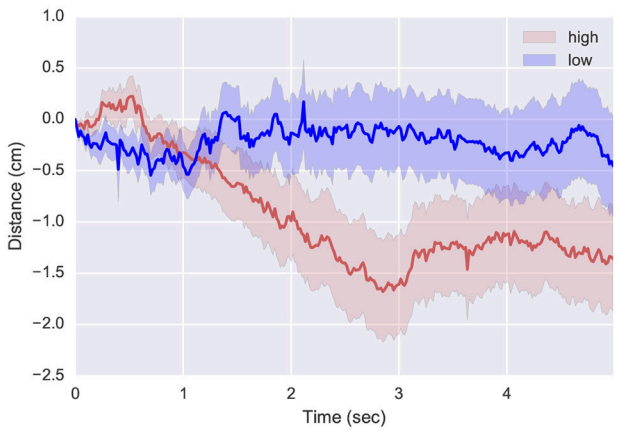

FIGURE 7 | The gaze distance over time for high and low interest responses. The shadow represent the 95% confidence intervals.

TABLE 5 | The list of nine GSR features.

Feature type	Description
Number of peaks	Number of peaks in resistance exceeding 100Ω
Amplitude of peaks	GSR peak amplitude from the saddle point preceding the peak
Rise time	The time it takes GSR to reach its peak from the saddle point in seconds
Statistical moments	Mean, first and third quartile & standard deviation (electrical resistance in Ω)
Trend	Intercept and slope for the linear trend

and complexity. Results were evaluated using Spearman rank correlation coefficients, due to the ordinal nature of the scores, and median absolute error (mAE). The random baseline for Spearman rank correlation is zero, and the random baseline with uniform distribution for normalized mAE ($\in [0, 1]$) is between 0.32 and 0.35. The results in images and micro-videos are reported separately.

We performed the same procedure for the regression on three modalities, namely, facial expression, eye gaze and GSR. We then performed the fusion using a weighted sum of the unimodal output (late fusion). The fusion weights were calculated from the training set as follows. After training the regression model we predicted the outcome from the training-set and fit a linear model that combined the modalities outcome to predict the target on the training-set. The regression evaluation results for images and micro-videos are given in **Table 6**. From the unimodal results, facial expressions are superior to the other modalities. GSR did not perform particularly well, due to the short duration of the trials. As expected, results on interest and curiosity are very similar. The interest detection results were superior for videos which might have more to offer in terms of interestingness or pleasantness.

Overall, coping potential and complexity were detected with higher accuracy compared to interest in image-viewing experiment. The superior performance for coping potential and complexity can be associated to their skewed distribution (see **Figure 4**).

We performed a one-tailed t-test to test the significance of the improvement in the multimodal detection from the best modality, i.e., facial expressions. Some of the multimodal results in the inter-participant cross-validation were superior in terms of MAE and for interest and curiosity in images we achieved significantly higher correlation for multimodal fusion.

We have also performed a one-participant-out cross validation. The results for micro-videos are reported in **Table 7**. The results for images were inferior and close to random ($\rho < 0.10$) which demonstrates that there were not enough similar patterns across participants in their responses to images. Participant independent results for micro-videos are inferior to

TABLE 6 | Multimodal and unimodal recognition model evaluation.

	Multimodal		Face		GSR		Gaze	
Scale	$\rho \uparrow$	**mAE ↓**	$\rho \uparrow$	**mAE ↓**	$\rho \uparrow$	**mAE ↓**	$\rho \uparrow$	**mAE ↓**
IMAGES								
Interest	0.23 (0.08)*	0.27 (0.02)	0.19 (0.06)	0.28 (0.02)	0.14 (0.05)	0.28 (0.02)	0.15 (0.10)	0.29 (0.02)
Curiosity	0.26 (0.06)*	0.27 (0.02)	0.23 (0.06)	0.28 (0.02)	0.20 (0.06)	0.27 (0.02)	0.15 (0.07)	0.29 (0.02)
Coping potential	0.37 (0.07)	0.18 (0.01)	0.35 (0.06)	0.19 (0.01)	0.24 (0.07)	0.19 (0.01)	0.17 (0.08)	0.21 (0.01)
Novelty	0.32 (0.04)	0.28 (0.02)*	0.30 (0.09)	0.29 (0.02)	0.23 (0.06)	0.28 (0.02)	0.18 (0.07)	0.32 (0.02)
Complexity	0.39 (0.06)	0.21 (0.01)	0.39 (0.05)	0.23 (0.01)	0.29 (0.07)	0.23 (0.02)	0.20 (0.06)	0.26 (0.01)
MICRO-VIDEOS								
Interest	0.31 (0.09)	0.26 (0.03)	0.29 (0.10)	0.27 (0.03)	0.05 (0.11)	0.29 (0.03)	0.27 (0.08)	0.27 (0.03)
Curiosity	0.35 (0.11)	0.28 (0.03)*	0.33 (0.11)	0.29 (0.03)	0.03 (0.09)	0.32 (0.02)	0.27 (0.10)	0.30 (0.03)
Coping potential	0.33 (0.14)	0.21 (0.02)	0.28 (0.14)	0.22 (0.02)	0.08 (0.07)	0.24 (0.02)	0.24 (0.09)	0.23 (0.01)
Novelty	0.32 (0.10)	0.30 (0.02)*	0.28 (0.11)	0.32 (0.02)	0.14 (0.11)	0.33 (0.02)	0.24 (0.10)	0.32 (0.02)
Complexity	0.35 (0.09)	0.25 (0.02)	0.34 (0.07)	0.26 (0.02)	0.13 (0.10)	0.28 (0.02)	0.14 (0.12)	0.28 (0.01)

*ρ is Spearman's ranking correlation and MAE is the median absolute error (mAE$\in$[0,1]). The numbers in parentheses are the standard deviation. *implies significantly higher than the best modality (face) with ($p < 0.05$).*

TABLE 7 | Participant independent multimodal and unimodal recognition model evaluation for micro-videos.

	Multimodal		Face		GSR		Gaze	
Scale	$\rho \uparrow$	**mAE ↓**	$\rho \uparrow$	**mAE ↓**	$\rho \uparrow$	**mAE ↓**	$\rho \uparrow$	**mAE ↓**
Interest	0.30 (0.19)	0.26 (0.09)	0.27 (0.18)	0.27 (0.03)	0.02 (0.18)	0.30 (0.10)	0.29 (0.20)	0.26 (0.10)
Curiosity	0.31 (0.15)	0.29 (0.09)	0.27 (0.16)	0.30 (0.09)	–0.04 (0.15)	0.34 (0.10)	0.29 (0.18)	0.30 (0.09)
Coping potential	0.26 (0.15)	0.22 (0.05)	0.19 (0.12)	0.23 (0.05)	0.01 (0.15)	0.24 (0.05)	0.26 (0.18)	0.22 (0.05)
Novelty	0.23 (0.18)	0.32 (0.08)	0.16 (0.17)	0.33 (0.09)	0.02 (0.17)	0.34 (0.08)	0.26 (0.20)	0.32 (0.09)
Complexity	0.15 (0.14)	0.28 (0.08)	0.08 (0.14)	0.29 (0.08)	0.01 (0.17)	0.29 (0.08)	0.17 (0.14)	0.28 (0.08)

ρ is Spearman's ranking correlation and MAE is the median absolute error (mAE$\in$[0,1]). The numbers in parentheses are the standard deviation.

the participant-dependent ones, however they remain significant. The results on recognizing complexity had the largest drop in performance between participant-dependent and independent evaluations.

5. DISCUSSIONS

The goal of our study was twofold. On the one hand, we wanted to contribute to the growing literature that is trying to differentiate among positive emotions and is looking into the specificities of each positive emotion (Campos et al., 2013; Shiota et al., 2017). For this objective, we chose interest for its many applications and relevance for our society (Clément and Dukes, 2013). On the other hand, we wanted to test whether a multimodal approach would benefit the automatic detection of interest and in consequence be the basis for future applications. Furthermore, we wanted to test a model that would be oriented toward the detection of both the emotion of interest and its constituting appraisals (Mortillaro et al., 2012).

In terms of expression, our results confirmed recent studies in that the experience of interest is generally related to a smiling behavior. This is an expressive feature that is common to positive emotions, and clearly it is not a feature that, alone, could be used to define the expression of interest. Conversely, interesting stimuli differed significantly from non interesting stimuli when the dynamics of the movement was considered: higher inner eyebrow (AU1) response in the first second for interesting stimuli and higher activation of the eye lid tightener (AU7) after the first second, likely as a consequence of immediate attention (AU1) and cognitive effort (AU7) (Silvia, 2008). In agreement with this finding, we found that novel and interesting stimuli are explored through longer saccades, an index of sustained attention. Interest was also associated with a different posture as indicated by the closer head distance measured by the eye tracker. All participants leaned toward the screen when a new image appeared on the screen (attention toward a novel stimulus), but only when the stimulus was interesting they maintained the posture and remain engaged; when the stimulus was not interesting they would go back to the resting position, distancing themselves from the screen. This is in line with recent work that points toward the important role of dynamics and bodily movements to define the nonverbal expression of positive emotions (Mortillaro et al., 2011; Dael et al., 2012; Dukes et al., 2017; Mortillaro and Dukes, 2018). It is important to note, that in terms of appraisals we found

correlations with Action Units that are in line with the most recent empirical evidence (Scherer et al., 2018).

In terms of automated recognition, interest and appraisal detection from GSR and gaze performed worse than facial expression for images. However, the fusion of face with any other modality slightly outperformed the unimodal regression results. Eye gaze results were also superior to the results from GSR and facial expression in response to micro-videos. A model, such as Papandreou et al. (2009), that can take the certainty of modalities in the multimodal fusion into account can possibly improve this fusion results.

The existing work on the automatic recognition of interest do not find facial expressions to be the most informative modality (Gatica-Perez et al., 2005; Schuller et al., 2009). For interest, unlike the so-called basic emotions (Ekman, 1993), there is no evidence that there is a unique and consistent facial expression. The findings on the temporal patterns of expressions and head pose motivated using machine learning methods that can learn temporal dependencies. We tried long-short-term memory recurrent neural networks (RNN) for this purpose but due to the small number of samples in this dataset, the RNN failed to achieve superior results. Recent advancement in computer vision is enabling vision-based gaze tracking (Wood et al., 2015) from single webcams. These findings motivates further work on detecting interest from audio- and vision-based methods that are also more practical in naturalistic or "in-the-wild" situations.

Our results are not at the same level as the ones reported by Schuller et al. (2009). However, there are a number of differences in the experiment and analysis. First, the protocol in Schuller et al. (2009) consist in an active social interaction whereas our recordings were done in non-social setting where participants are less expressive. Second, their ground-truth was generated by the third-person labelers which are more consistent compared to the self-reports with participant-dependent bias. Despite this relatively low performance, our results indicates that the detection of interest can profit of the integration of multiple modalities. If only one modality should be used–for technical constraints or real-world applications—facial expression is a good candidate. However, this should be better done taking into account the dynamics of the movements and if possible head movements as well.

We also tested an alternative approach to emotion recognition in this research. Based on the suggestions of Mortillaro et al. (Mortillaro et al., 2012; Scherer et al., 2018) we explored the possibility to automatically detect appraisals. We focused our attention on the appraisals that are most relevant for interest and achieved promising results. For images, coping, potential, novelty and complexity were better recognized than interest in all unimodal detection procedure and also in the multimodal approach. This pattern of results suggests to consider appraisals as building blocks for emotion expressions and use them as first-level target of emotion detection algorithms (Mortillaro et al., 2012; De Melo et al., 2014; Scherer et al., 2018). We did not find the same pattern of results for micro-videos, but this is mostly due to the better performance in detecting interest when participants watched micro-videos. One could speculate that when the stimuli are rich in information and people are expressive, interest can be directly recognized due to its strong association with positive expressions. On the contrary when the emotions are less intense and the expressions are very subtle, systems should probably target appraisals as these may be the best elements to differentiate between subtly different emotional states, like positive emotions (Mortillaro et al., 2011).

Previous work on automatic recognition of visual interestingness from the visual content (Gygli et al., 2013; Soleymani, 2015; Gygli and Soleymani, 2016) reported higher performance for detecting average interest, with correlation reaching 0.71 in Gygli et al. (2013) for images and 0.53 for micro-videos (Gygli and Soleymani, 2016). Further attempts at recognizing appraisals can be combined with the content analysis to detect interest from both expressions and the content. For example, if intrinsic pleasantness and aesthetics are related to interest in images, as is shown in Soleymani (2015) and Gygli and Soleymani (2016), the visual content can be analyzed or tagged on the degree of its pleasantness and aesthetics. The recognized appraisals, such as novelty, can be then used with intrinsic pleasantness for interest detection.

In this work, we found an association between interest and smile. However, it is important to note that we only studied the behavior displayed in reaction to stimuli that were rated as interesting or not. We did not compare the expression of interest with the expression of other emotions. In the latter case, we would likely find smiling in most positive emotions (Ekman, 1992, 1993; Campos et al., 2013). Smiling is one indicator of interest in our experimental framework, but is not the only one nor the most defining. Smiling is not a simple behavior that reflects one genuine emotion, but a powerful behavior that can have multiple social and emotional functions (e.g., Rychlowska et al., 2017). For example, smiling can be the expression of other positive emotions, or be used to foster bonding or dominance. It is for this reason that we suggest using an appraisal-based approach to emotion recognition, that allows a greater flexibility than the traditional pattern-matching approach. In our study, we related smiling to the appraisal of intrinsic pleasantness, and we cannot fully exclude the possibility that other positive emotions co-occurred with the experience of interest. Researchers who study interest should be aware of this special role of smiling and consider all the constraints and implications of their research paradigm and applications in their models.

In this work, the content was limited to visual content with no personal connection to the participants. However, in practice the relevance of the content or personal connection to the user is an important factor in determining its interestingness. A grainy picture of a loved one might be more interesting than a sharp and aesthetically pleasing image of a random scene. This limitation can be addressed in the future by adding personally relevant and irrelevant content to assess the appraisal of relevance.

Another limitation of this work is that the participants did not have any specific task or goal, and the person was passively looking at the stimuli. This passive role combined with the absence of social interaction limited the number of expressions that participants displayed.

In our study we used only 80 pictures and 40 short videos. Having a larger dataset will be also beneficial for training models that can learn temporal dependencies, e.g., recurrent neural networks.

Finally, micro-videos elicited more consistent behavioral patterns across participants, as is observable in the participant-independent results. We believe that the still images could not elicit emotions and reactions as strong as those elicited by moving pictures and therefore we suggest using videos in future work.

The current work addressed the automatic detection of short term and episodic interest also known as situational interest. That is different than what most current recommender systems do, that is identify longer term personal interest based on content analysis. Obviously, using behavioral signals such as facial expression to detect situational interest requires capturing facial images and we should be aware that users might find that intrusive, for at least two reasons. First, users might not want to share information that would make them identifiable with a system. Second, one might not necessarily want to share his/her inner state such as interest in a given content. Deploying such systems should be only done with the full informed consent of its users and the users should have full control over how and where the data can be used. Such systems should be designed not to transfer or store identifiable information, in this case facial images. One existing solution is to execute facial tracking on users device and only transfer or store the analysis outcome.

6. CONCLUSIONS

An automatic approach for detecting interest has application in different domains. For example, visual interest can be used to re-rank images in a recommender or retrieval systems (Walber et al., 2014). Similar methods can also have applications in online education (D'Mello et al., 2007) and marketing (Kurdyukova et al., 2012).

In this work, we conducted an experiment with the goal of assessing interest, curiosity and their relevant appraisals while participants watched visual stimuli. We found temporal patterns of facial expression, posture and head pose that are related to interest. Analysis of GSR demonstrated that interest is likely associated with higher arousal. Eye gaze patterns of the users with a higher level of interest contain longer saccades which might be associated with the action tendency of exploration. Analysis of facial expressions shows that interest is related to eye opening and smile, which are signs of novelty and pleasantness. The correlation between smile and interest is in agreement with our previous findings in a similar context which showed positive correlation between pleasantness and interest (Mortillaro et al., 2011; Soleymani, 2015).

Our results about the relationship between appraisals and Facial Action Units are generally in line with the most recent empirical findings (Scherer et al., 2018). Our study also showed that appraisals are related to eye gaze (longer saccades for stimuli appraised as novel) and that electrodermal responses or fluctuations are also associated with higher level of interest. The results of the detection algorithms reflect these outcomes, i.e., appraisals were better detected than interest in all modalities and in the multimodal approach. All in all, we suggest that future attempts at detecting subtle or non-basic emotions may focus on detecting appraisals and adopt a multimodal approach. The appraisals should then be used as the building blocks for detecting interest or other emotions that do not have a prototypical expression.

AUTHOR CONTRIBUTIONS

MS designed the experiment, and collected the data and performed the quantitative analysis. MM contributed to the formulation, discussions and writing of the manuscript.

FUNDING

The work of MS was supported by his Ambizione grant from the Swiss National Science Foundation. The work of MM was supported by the National Centre of Competence in Research (NCCR) Affective Sciences: Emotion in individual Behaviour and Social Processes, financed by the Swiss National Science Foundation [grant number: SNSF, 51NF40-104897], and hosted by the University of Geneva.

ACKNOWLEDGMENTS

We would like to thank Danny Dukes for discussions on the psychology of interest. We thank David Sander for his kind assistance during the process of the ethical review of the protocol. We acknowledge the support from the Fondation Campus Biotech Genève for their generous support with the experimental facilities and the computer science department, University of Geneva, for providing the material support necessary for running the experiments.

REFERENCES

Bakhshi, S., Shamma, D. A., Kennedy, L., Song, Y., de Juan, P., and Kaye, J. J. (2016). "Fast, cheap, and good: why animated GIFs Engage Us," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, CA), 575–586. doi: 10.1145/2858036.2858532

Baltrusaitis, T., Mahmoud, M., and Robinson, P. (2015). "Cross-dataset learning and person-specific normalisation for automatic Action Unit detection," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Piscataway, NJ), 1–6. doi: 10.1109/FG.2015.7284869

Baltrusaitis, T., Robinson, P., and Morency, L.-P. (2016). "OpenFace: An open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (Piscataway, NJ), 1–10. doi: 10.1109/WACV.2016.7477553

Bixler, R., and D'Mello, S. (2013). "Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits," in *Proceedings*

of the 2013 International Conference on Intelligent User Interfaces, IUI '13 (New York, NY: ACM), 225–234. doi: 10.1145/2449396.2449426

Blanchard, N., Bixler, R., Joyce, T., and D'Mello, S. (2014). "Automated Physiological-Based Detection of Mind Wandering during Learning," in *12th International Conference, ITS 2014, Proceedings* (Honolulu, HI: Springer International Publishing), 55–60.

Calvo, R. A., and D'Mello, S. (2010). Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affect. Comput.* 1, 18–37. doi: 10.1109/T-AFFC.2010.1

Campos, B., Shiota, M. N., Keltner, D., Gonzaga, G. C., and Goetz, J. L. (2013). What is shared, what is different? Core relational themes and expressive displays of eight positive emotions. *Cogn. Emot.* 27, 37–52. doi: 10.1080/02699931.2012.683852

Camras, L. A., Bakeman, R., Chen, Y., Norris, K., and Cain, T. R. (2006). Culture, ethnicity, and children's facial expressions: a study of European American, Mainland Chinese, Chinese American, and adopted Chinese girls. *Emotion* 6, 103–114. doi: 10.1037/1528-3542.6.1.103

Clément, F., and Dukes, D. (2013). The role of interest in the transmission of social values. *Front. Psychol.* 4:349. doi: 10.3389/fpsyg.2013.00349

Dael, N., Mortillaro, M., and Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion* 12, 1085–1101. doi: 10.1037/a0025737

De Melo, C. M., Carnevale, P. J., Read, S. J., and Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *J. Pers. Soc. Psychol.* 106, 73–88. doi: 10.1037/a0034251

D'Mello, S., Olney, A., Williams, C., and Hays, P. (2012). Gaze tutor: a gaze-reactive intelligent tutoring system. *Int. J. Hum. Comput. Stud.* 70, 377–398. doi: 10.1016/j.ijhcs.2012.01.004

D'Mello, S., Picard, R. W., and Graesser, A. (2007). Toward an affect-sensitive autotutor. *IEEE Intell. Syst.* 22, 53–61. doi: 10.1109/MIS.2007.79

Dukes, D., Clément, F., Audrin, C., and Mortillaro, M. (2017). Looking beyond the static face in emotion recognition: the informative case of interest. *Vis. Cogn.* 25, 575–588. doi: 10.1080/13506285.2017.1341441

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200. doi: 10.1080/02699939208411068

Ekman, P. (1993). Facial expression and emotion. *Am. Psychol.* 48, 384–392. doi: 10.1037/0003-066X.48.4.384

Ekman, P., and Friesen, W.V. (1978). *Manual for the Facial Action Coding System,* Palo Alto, CA: Consulting Psychologists Press.

Ellsworth, P. C., and Smith, C. A. (1988). Shades of Joy: patterns of Appraisal Differentiating Pleasant Emotions. *Cogn. Emot.* 2, 301–331. doi: 10.1080/02699938808412702

Gatica-Perez, D., McCowan, I., Zhang, D., and Bengio, S. (2005). "Detecting group interest-level in meetings," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Piscataway, NJ), 489–492. doi: 10.1109/ICASSP.2005.1415157

Gonzaga, G. C., Turner, R. a., Keltner, D., Campos, B., and Altemus, M. (2006). Romantic love and sexual desire in close relationships. *Emotion (Washington, D.C.)* 6, 163–179. doi: 10.1037/1528-3542.6.2.163

Hatice, G., Björn, S., Maja, P., and Roddy, C. (2011). "Emotion representation, analysis and synthesis in continuous space: A survey," in *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference* (Piscataway, NJ), 827–834. doi: 10.1109/FG.2011.5771357

Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., and Van Gool, L. (2013). "The interestingness of images," in *The IEEE International Conference on Computer Vision (ICCV)* (Piscataway, NJ). doi: 10.1109/ICCV.2013.205

Gygli, M., Grabner, H., and Van Gool, L. (2015). "Video summarization by learning submodular mixtures of objectives," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Piscataway, NJ). doi: 10.1109/CVPR.2015.7298928

Gygli, M., and Soleymani, M. (2016). "Analyzing and predicting gif interestingness," in *Proceedings of the 2016 ACM on Multimedia Conference, MM '16* (New York, NY: ACM), 122–126. doi: 10.1145/2964284.2967195

Gygli, M., Song, Y., and Cao, L. (2016). "Video2GIF: automatic generation of animated GIFs from video," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV). doi: 10.1109/CVPR.2016.114

Jennifer, H., and Rosalind, P. (2000). "SmartCar: detecting driver stress," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (Barcelona), 218–221. doi: 10.1109/ICPR.2000.902898

Izard, C. E. (2009). Emotion theory and research: highlights, unanswered questions, and emerging issues. *Annu. Rev. Psychol.* 60, 1–25. doi: 10.1146/annurev.psych.60.110707.163539

Jaques, N., Conati, C., Harley, J. M., and Azevedo, R. (2014). "Predicting affect from gaze data during interaction with an intelligent tutoring system," in *12th International Conference, ITS 2014, Proceedings, Volume 8474 LNCS* (Honolulu, HI: Springer International Publishing), 29–38. doi: 10.1007/978-3-319-07221-0_4

Kapoor, A., and Picard, R. W. (2005). "Multimodal affect recognition in learning environments," in *Proceedings of the 13th annual ACM international conference on Multimedia - MULTIMEDIA '05* (New York, NY: ACM Press), 677. doi: 10.1145/1101149.1101300

Kapoor, A., Picard, R. W., and Ivanov, Y. (2004). "Probabilistic combination of multiple modalities to detect interest," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, Vol. 3, (Piscataway, NJ), 969–972. doi: 10.1109/ICPR.2004.1334690

Kim, J., and André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Trans. Patt. Anal. Mach. Intell.* 30, 2067–2083. doi: 10.1109/TPAMI.2008.26

Kleinsmith, A., and Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: a survey. *IEEE Trans. Affect. Comput.* 4, 15–33. doi: 10.1109/T-AFFC.2012.16

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* 84, 394–421. doi: 10.1016/j.biopsycho.2010.03.010

Kurdyukova, E., Hammer, S., and André, E. (2012). "Personalization of content on public displays driven by the recognition of group context," in *Ambient Intelligence, Vol. 7683,* eds F. Paternò, B. de Ruyter, P. Markopoulos, C. Santoro, E. van Loenen, and K. Luyten (Berlin; Heidelberg: Springer), 272–287.

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

Matias, R., Cohn, J. F., and Ross, S. (1989). A comparison of two systems that code infant affective expression. *Dev. Psychol.* 25, 483–489. doi: 10.1037/0012-1649.25.4.483

Mortillaro, M., and Dukes, D. (2018). Jumping for joy: the importance of the body and dynamics in the expression and recognition of positive emotions. *Front. Psychol.* 9:763. doi: 10.3389/fpsyg.2018.00763

Mortillaro, M., Mehu, M., and Scherer, K. R. (2011). Subtly different positive emotions can be distinguished by their facial expressions. *Soc. Psychol. Pers. Sci.* 2, 262–271. doi: 10.1177/1948550610389080

Mortillaro, M., Meuleman, B., and Scherer, K. R. (2012). Advocating a componential appraisal model to guide emotion recognition. *Int. J. Synthet. Emot.* 3, 18–32. doi: 10.4018/jse.2012010102

Papandreou, G., Katsamanis, A., Pitsikalis, V., and Maragos, P. (2009). Adaptive multimodal fusion by uncertainty compensation with application to audiovisual speech recognition. *IEEE Trans. Audio Speech Lang. Process.* 17, 423–435. doi: 10.1109/TASL.2008.2011515

Reeve, J. (1993). The face of interest. *Motivat. Emot.* 17, 353–375. doi: 10.1007/BF00992325

Reeve, J., and Nix, G. (1997). Expressing intrinsic motivation through acts of exploration and facial displays of interest. *Motivat. Emot.* 21, 237–250. doi: 10.1023/A:1024470213500

Rychlowska, M., Jack, R. E., Garrod, O. G., Schyns, P. G., Martin, J. D., and Niedenthal, P. M. (2017). Functional smiles: tools for love, sympathy, and war. *Psychol. Sci.* 28, 1259–1270. doi: 10.1177/0956797617706082

Sabourin, J., Mott, B., and Lester, J. C. (2011). "Modeling learner affect with theoretically grounded dynamic Bayesian networks," in *Affective Computing and Intelligent Interaction* (Memphis, TN: Springer Berlin Heidelberg), 286–295. doi: 10.1007/978-3-642-24600-5_32

Sariyanidi, E., Gunes, H., and Cavallaro, A. (2015). Automatic analysis of facial affect: a survey of registration, representation, and recognition. *IEEE Trans. Patt. Anal. Mach. Intell.* 37, 1113–1133. doi: 10.1109/TPAMI.2014.2366127

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Soc. Sci. Inform.* 44, 695–729. doi: 10.1177/0539018405058216

Scherer, K. R. (2009). The dynamic architecture of emotion: evidence for the component process model. *Cogn. Emot.* 23, 1307–1351. doi: 10.1080/02699930902928969

Scherer, K. R., Mortillaro, M., Rotondi, I., Sergi, I., and Trznadel, S. (2018). Appraisal-driven facial actions as building blocks for emotion inference. *J. Pers. Soc. Psychol.* 114, 358–379. doi: 10.1037/pspa0000107

Scherer, K. R., Zentner, M. R., and Stern, D. (2004). Beyond surprise: the puzzle of infants' expressive reactions to expectancy violation. *Emotion* 4, 389–402. doi: 10.1037/1528-3542.4.4.389

Schuller, B., Müller, R., Eyben, F., Gast, J., Hörnler, B., Wöllmer, M., et al. (2009). Being bored? Recognising natural interest by extensive audiovisual integration for real-life application. *Image Vis. Comput.* 27, 1760–1774. doi: 10.1016/j.imavis.2009.02.013

Shiota, M. N., Campos, B., Oveis, C., Hertenstein, M. J., Simon-Thomas, E., and Keltner, D. (2017). Beyond happiness: building a science of discrete positive emotions. *Amer. Psychol.* 72, 617–643. doi: 10.1037/a0040456

Silvia, P. J. (2005). What is interesting? Exploring the appraisal structure of interest. *Emotion* 5, 89–102. doi: 10.1037/1528-3542.5.1.89

Silvia, P. J. (2006). *Exploring the Psychology of Interest*. New York, NY: Oxford University Press.

Silvia, P. J. (2008). Interest The Curious Emotion. *Curr. Direct. Psychol. Sci.* 17, 57–60. doi: 10.1111/j.1467-8721.2008.00548.x

Silvia, P. J., Henson, R. A., and Templin, J. L. (2009). Are the sources of interest the same for everyone? Using multilevel mixture models to explore individual differences in appraisal structures. *Cogn. Emot.* 23, 1389–1406. doi: 10.1080/02699930902850528

Soleymani, M. (2015). "The quest for visual interest," in *Proceedings of the 23rd Annual ACM Conference on Multimedia*, (New York, NY), 919–922.

Soleymani, M., Pantic, M., and Pun, T. (2012). Multimodal emotion recognition in response to videos. *IEEE Trans. Affect. Comput.* 3, 211–223. doi: 10.1109/T-AFFC.2011.37

Soleymani, M., Villaro-Dixon, F., Pun, T., and Chanel, G. (2017). Toolbox for emotional feature extraction from physiological signals (teap). *Front. ICT* 4:1. doi: 10.3389/fict.2017.00001

Tracy, J. L., and Robins, R. W. (2004). Show your pride: evidence for a discrete emotion expression. *Psychol. Sci.* 15, 194–197. doi: 10.1111/j.0956-7976.2004.01503008.x

Walber, T. C., Scherp, A., and Staab, S. (2014). "Smart photo selection: interpret gaze as personal interest," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14* (New York, NY: ACM), 2065–2074. doi: 10.1145/2556288.2557025

Weninger, F., Wöllmer, M., and Schuller, B. (2015). "Emotion recognition in naturalistic speech and language-A survey," in *Emotion Recognition* (Hoboken, NJ: John Wiley & Sons, Inc.), 237–267.

Wood, E., Baltruaitis, T., Zhang, X., Sugano, Y., Robinson, P., and Bulling, A. (2015). "Rendering of eyes for eye-shape registration and gaze estimation," in *2015 IEEE International Conference on Computer Vision (ICCV)* (Piscataway, NJ), 3756–3764. doi: 10.1109/ICCV.2015.428

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.