University of Glasgow

Gender and Emotion Recognition with Implicit User Signals

Maneesh Bilalpur
International Institute of Information
Technology
Hyderabad, India
maneesh.bilalpur@research.iiit.ac.in

Seyed Mostafa Kia
Donders Institute, Radboud
University
Nijmegen, Netherlands
s.kia@donders.ru.nl

Manisha Chawla
Centre for Cognitive Science, Indian
Institute of Technology
Gandhinagar, India
manisha.chawla@iitgn.ac.in

Tat-Seng Chua
School of Computing, National
University of Singapore
Singapore
chuats@comp.nus.edu.sg

Ramanathan Subramanian
University of Glasgow & Advanced
Digital Sciences Center
Singapore
ramanathan.subramanian@glasgow.
ac.uk

ABSTRACT

We examine the utility of ***implicit user behavioral signals*** captured using low-cost, off-the-shelf devices for anonymous gender and emotion recognition. A user study designed to examine male and female sensitivity to facial emotions confirms that females recognize (especially negative) emotions *quicker* and *more accurately* than men, mirroring prior findings. Implicit viewer responses in the form of ***EEG brain signals*** and ***eye movements*** are then examined for existence of (a) emotion and gender-specific patterns from *event-related potentials* (ERPs) and *fixation distributions* and (b) emotion and gender discriminability. Experiments reveal that (i) Gender and emotion-specific differences are observable from ERPs, (ii) multiple similarities exist between *explicit* responses gathered from users and their *implicit* behavioral signals, and (iii) Significantly above-chance ($\approx$70%) gender recognition is achievable on comparing emotion-specific EEG responses– gender differences are encoded best for *anger* and *disgust*. Also, fairly modest *valence* (positive vs negative emotion) recognition is achieved with EEG and eye-based features.

CCS CONCEPTS

•**Human-centered computing** → HCI theory, concepts and models; *User centered design;*

KEYWORDS

Gender differences; Facial emotion processing; Gender and Emotion Recognition; Implicit behavioral signals; EEG; Eye movements

1 INTRODUCTION

Gender HCI [54] and **Affective HCI** [39] have evolved as critical HCI sub-fields due to widespread acknowledgment of the fact that computers need to *appreciate* and *adapt to* the user's gender and emotional state. The ability to identify user demographics including gender and emotion can benefit interactive and gaming systems in terms of a) visual and interface design [10, 37], (b) game and product recommendation (via ads) [17, 57], and (c) provision of appropriate motivation and feedback for optimizing user experience [43]. Contemporary gender recognition (GR) and emotion recognition (ER) systems primarily work with facial [18, 35] or

speech [25, 26] cues; however, face and speech are *biometrics* that encode an individual's identity, and pose grave privacy concerns as they can be recorded without the user's knowledge [1].

This work examines GR and ER from ***implicit user behavioral signals***, in the form of *EEG brain signals* and *eye movements*. The conveyance of implicit behavioral signals is hidden from the outside world, and they cannot be recorded without express user cooperation making them privacy compliant [9]. Also, behavioral signals such as EEG and eye movements are primarily *anonymous* as little is known regarding their uniqueness to a person's identity [56].

Specifically, we attempt GR and ER using signals captured by commercial, off-the-shelf devices which are minimally intrusive, affordable and popularly used in gaming as input or feedback modalities [2, 28, 50]. The *Emotiv* EEG wireless headset consists of 14 dry (plus two reference) electrodes having a configuration as shown in Fig. 1. While being lightweight, wearable and easy-to-use, neuro-analysis with Emotiv can be challenging due to relatively poor signal quality. Likewise, the *EyeTribe* is a low-cost eye-tracker whose suitability for research purposes has been evaluated and endorsed [11]. We show how relevant gender and emotion-specific information is captured by these low-cost devices via examination of event-related potential (ERP) and fixation distribution patterns, and also through recognition experiments.
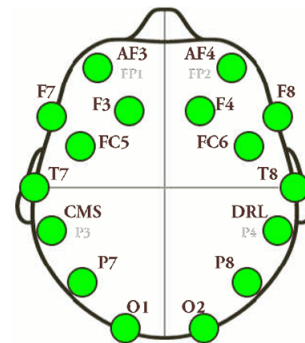

Figure 1: Emotiv electrode configuration.

To capture gender-based differences, we designed a facial emotion recognition experiment as men and women have been known to respond differently to affective information [7, 14, 32, 43]. Our

study performed with 28 viewers (14 male) confirms that women are superior at facial ER, mirroring prior findings. Hypothesizing that enhanced female emotional sensitivity should reflect via their implicit responses, we examined EEG signals and found that (1) Stronger ERPs are observed for females while processing negative facial emotions, (2) Better ER is achieved with female EEG responses and eye movement features, and (3) Emotion-specific gender differences manifest better than emotion-agnostic differences.

Overall, our work makes the following research contributions: (a) While prior works have identified gender differences in emotional behavior, this is one of the first works to expressly perform GR and ER from implicit behavioral signals; (b) Apart from recognition experiments, we show that the employed devices capture meaningful information, as typified by gender and emotion-specific event related potentials (ERPs); (c) The use of minimally intrusive, off-the-shelf and low-cost devices extends the ecological validity of our findings, and the feasibility of employing a similar framework for large-scale interactive user profiling.

The paper is organized as follows. Section 2 reviews related work. Section 3 describes the experimental settings and protocol adopted to examine gender differences. Section 4 examines *explicit* user responses in terms of response times and recognition rates, while Section 5 discusses *implicit* user behavior based on EEG and gazing patterns while also presenting GR and ER experiments. Section 6 discusses the salient aspects of this work and concludes the paper.

2 RELATED WORK

To position our work with respect to the literature, this section reviews related work on (a) user-centered ER, and (b) gender differences in emotional face processing.

2.1 User-centered ER

Emotions evoked by multimedia content have been modeled via *content-centered* as well as *user-centered* approaches. Content-centered approaches typically attempt to find emotional correlates from content-based features [16, 44, 53], while user-centered methods monitor user behavioral cues (eye movements, EEG signals, *etc*) to recognize the evoked emotion. As emotion represents a subjective human feeling, many user-centered approaches have attempted to model emotions by examining *explicit* and *implicit* user behavioral cues. Conspicuous facial cues are studied in [18] to detect multimedia highlights, while implicit physiological measurements are employed to model emotions induced by music and movie scenes respectively in [20] and [3]. EEG and eye movements are two popular implicit modalities employed for ER, and many works have predicted affect with a combination of both [7, 30, 58], or exclusively using either signal [3, 19, 27, 40, 46, 47, 59].

Valence (positive vs negative emotion) recognition using eye-based features is proposed in [40]. ER with EEG and pupillary responses acquired from five users is discussed in [58]. A deep unsupervised method for ER from raw EEG data is proposed in [27], and its effectiveness is shown to be comparable to manually extracted features. Differential entropy (DE) features from EEG data are extracted to train an integrated deep belief network plus hidden Markov model for ER in [59]. A DAE (differential auto-encoder) that learns shared representations from EEG and eye-based features is

proposed for valence recognition in [30]. Almost all of these works employ lab-grade eye-trackers and EEG sensors which are bulky and intrusive, and therefore preclude naturalistic user behavior.

2.2 Gender Differences in Emotion Recognition

As facial emotions denote important non-verbal communication cues during social interactions, many social psychology studies have examined human recognition of emotional faces. Certain facial features are found to encode emotional cues better than others. When viewing a human face, most attention is drawn to the eyes, nose and mouth regions [45, 52]. Authors of [48] observe that visual attention is localized around the eyes for mildly emotive faces, but the nose and mouth regions also attract substantial eye fixations in highly emotive faces. A recent eye tracking study [42] notes that distinct eye fixation patterns emerge for different facial emotions. The mouth is most informative for the *joy* and *disgust* emotions, whereas eyes mainly encode information relating to *sadness*, *fear*, *anger* and *shame*. In a similar study [4], more fixations are noted on the upper face half for *anger* as compared to *disgust*, while no differences are observed on lower face half for the two emotions. However, humans can find it difficult to distinguish between some facial emotions that tend to have similar characteristics– examples are the high overlap rate between the *fear−surprise* and *anger−disgust* emotion pairs [12, 45].

Many studies have also identified gender differences during facial emotion processing. Females are generally found to be better at ER than males, irrespective of age [49]. Other studies examining the role of facial movements in ER [6, 14, 32] also note that females recognize facial emotions more accurately than males, even when only partial information is available. Some evidence also points to the fact that females are faster at ER than males [15, 41].

Differences in gaze patterns and neural activations have been found between males and females while viewing emotional faces. Women's tendency to fixate on the eyes positively correlates with their ER capabilities, while men tend to look at the mouth for emotional cues [15, 49]. Likewise, there exists evidence in terms of EEG ERPs that negative facial emotions are processed differently and rapidly by women, and do not necessarily entail selective attention towards emotional cues [29, 60].

An exhaustive review of GR methodologies is presented in [55], and the authors evaluate GR methods across different communities using metrics like *universality, distinctiveness, permanence* and *collectability*. While crediting bio-signals like EEG and ECG for their accuracy and trustworthiness, this work also highlights the invasiveness of bio-sensors. The sensors used in this work are nevertheless minimally intrusive and non-invasive, thereby ensuring naturalistic user experience, while also recording meaningful emotion and gender-related information. Among the few works to attempt GR from user-centric cues, EEG and speech features are proposed for age and gender recognition in [36].

2.3 Analysis of Related Work

A close examination of related work reveals that (1) Many have attempted ER from user-centered cues, both conspicuous and implicit, and a number of works have also tried to isolate gender-specific

differences based on gaze patterns and neural activations; nevertheless, very few works have that expressly attempted GR based on implicit user cues. Differently, this work employs implicit signals for GR and ER with a number of classifiers and reliable detection of both gender and valence (AUC score of >0.6) is achieved via our proposed methodology; (2) Another salient aspect of our approach is the use of low cost, off-the-shelf and easy-to-use sensors for acquiring implicit signals, which suffer from poor signal quality while enabling natural user behavior. We show how these sensors nevertheless capture meaningful information via the analysis of ERPs and fixation distribution patterns; (3) In contrast to most prior works have either analyzed (a) explicit user responses in terms of reaction times and recognition rates, or (b) implicit behavioral patterns to discover gender differences in ER, or (c) employed EEG and eye-based features for GR and ER without validation of the corresponding signals, our work touches upon all of these aspects. We show how explicit behavioral gender differences noted for negative emotions also reflect via EEG ERPs and fixation distributions, and GR performance. The multiple similarities among explicit and implicit behavioral patterns, along with the use of low-cost sensors validate the findings in this work and extend its ecological validity. The next section describes the experimental set up and protocol employed in our study.

3 MATERIALS AND METHODS

3.1 Experimental Overview

We investigated gender differences during visual emotional face processing via explicit recognition rates and implicit EEG plus eye movement patterns. While emotion elicitation can and has been achieved via the presentation of emotional faces to viewers (*e.g.*, IAPS pictures [23]), our objective was to study *facial emotion recognition* rather than *affect elicitation*. Specifically, we wanted to validate prior findings regarding the enhanced sensitivity of females to facial emotions. Experimental details are as follows:

3.1.1 Participants: 28 subjects from different nationalities (14 male, age 26.1 ± 7.3 and 14 female, age 25.5 ± 6) took part in our study. All of them were university engineering graduate and undergraduate students, and therefore likely possessed similar levels of intellect and social intelligence. All subjects had normal or corrected vision and provided informed consent for participation.

3.1.2 Stimuli: We used the facial emotions of 24 models (12 male, 12 female) from the Radboud Faces Database (RaFD) [24]. RaFD contains facial emotions of 49 models which are rated for clarity, genuineness and intensity. 24 models were chosen such that their Ekman facial emotions[1] were roughly matched with respect to the above ratings. We also *morphed* the emotive faces from *neutral* (0% intensity) to *maximum* (100% intensity) to generate intermediate morphs in steps of 5%. Since viewers had difficulty recognizing very low intensity morphs, we only used 25-100% morphs in our experiments. To examine the impact of morph intensity on ER performance, we divided the morphed emotive faces into *low-intensity* (LI, 25–50%) and *high-intensity* (HI, 55–100%) morphs (see Fig. 2). This resulted in a dataset comprising 864 LI and 1440 HI emotional faces. All original and morphed faces were 361×451 pixels in size,

[1]Anger (A), Disgust (D), Fear (F), Happy (H), Sad (Sa) and Surprise (Su).
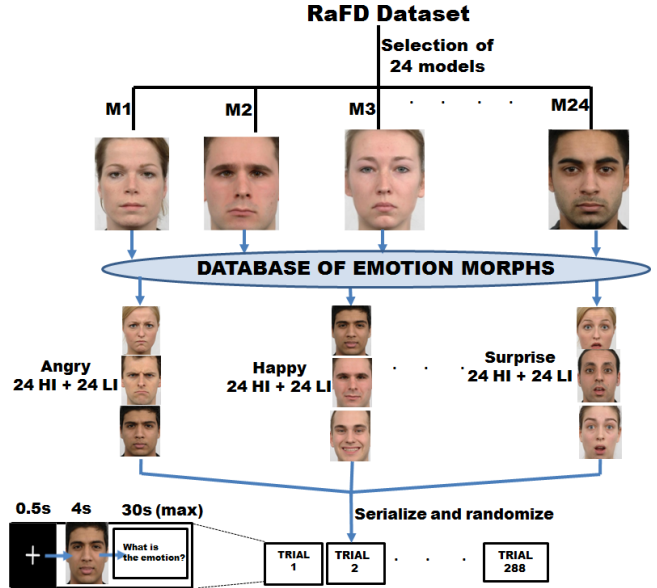

Figure 2: Experimental Protocol: Timeline for each trial is shown on the bottom-left inset. Best-viewed under zoom.

translating to a visual angle of 9.1° and 11.4° about x and y at 60 cm screen distance.

3.1.3 Protocol: The experimental protocol was developed using Matlab *Psychtoolbox* [8], and is outlined in Fig. 2. Corresponding to each model and emotion, we chose one HI and LI morph at random. This generated a total of 288 face images (2 morphs/emotion × 6 emotions × 24 models) for presentation to each viewer, and these images were displayed in random order. In each trial (left bottom inset in Fig. 2), an emotive face was displayed for 4s preceded by a fixation cross for 500 ms. Upon image presentation, the viewer had a time-limit of 30s to make *one* out of *seven* choices concerning the facial emotion (6 emotions plus neutral) via a radio button. Neutral faces were only used for morphing and were not presented during the experiment. Viewer EEG responses were recorded via the 14-channel *Emotiv epoc* device, while their eye-movements were captured via the *Eyetribe* eye-tracker. To minimize calibration errors, the experiment was split into 4 parts comprising 72 trials each, and took 90 minutes to complete.

4 USER RESPONSE ANALYSIS

We now examine user behavioral data and compare male and female performance in the facial emotion recognition task based on (i) response times (RTs) and (ii) recognition rates (RRs).

4.1 Response time (RT)

The overall viewer RT was 1.5±0.07 s, implying that users judged facial emotions fairly instantaneously. Table 1 summarizes results from a 3-way ANOVA on RTs, with ***morph intensity*** (MI) and *facial **emotion*** as within-subject factors, and ***gender*** as the between-subject factor. ANOVA revealed the main effect of MI with generally faster RTs noted for HI morphs (see Fig. 3). The presented

Table 1: ANOVA summary for behavioral measures. df and Sig respectively denote degrees of freedom and significance level.

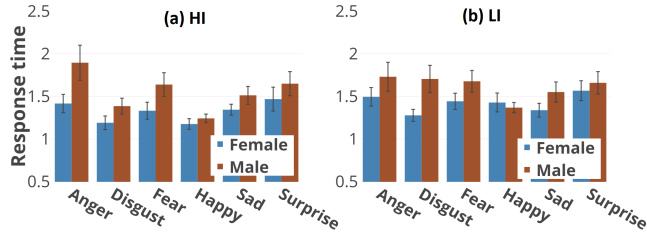| Response	Predictor	Morph Intensity	Emotion	Gender	MI*Emotion	MI*Gender	Error	Total
	df	1	5	1	5	5	32	160
RT	**F**	5.179	7.699	3.411	2.901	3.327		
	Sig	$p < 0.05$	$p < 0.001$	$p = 0.074$	$p < 0.05$	$p < 0.05$		
RR	**F**	531.1	59.83	7.37		7.359		
	Sig	$p < 0.001$	$p < 0.001$	$p < 0.05$		$p < 0.001$		

Figure 3: Emotion-wise mean RTs of males and females considering HI and LI morphs. Error bars denote unit standard error (best-viewed in color).

Figure 4: Male and female RRs for HI morphs, and LI morphs.

facial motion also impacted RTs, and pairwise comparisons revealed that RT for *happy* (1.3 ± 0.9 s) was significantly faster than for *anger*, *fear*, *sad* and *surprise*. ANOVA also revealed a marginal gender effect with females (1.4 ± 0.1 s) generally responding faster than males (1.6 ± 0.1 s). Significant interaction effects were also noted between (a) morph intensity and emotion, and (b) gender and emotion. As evident from Fig. 3, females were significantly faster ($F(3, 96) = 3.33, p < 0.05$) at judging negative (*anger*, *disgust*, *fear* and *sad*) emotions.

4.2 Recognition rates (RRs)

A 64% RR was achieved overall, suggesting that facial emotions were recognized well above chance level but well below ceiling. ANOVA summary in Table 1 reveals the main effect of morph intensity, emotion and gender. Females were more accurate than males at recognizing negative emotions, even from LI morphs (Fig. 4(a,b)). HI morphs were recognized more accurately (mean RR = 78.7) than LI morphs (mean RR = 49.1). Also, *happy* faces were recognized most accurately, and *fearful* faces least accurately. We also found a significant interaction effect between morph intensity and emotion. Overall, (a) HI emotion morphs were recognized way more accurately than LI morphs, (b) females recognized negative emotions better than males and (c) *Happy* was the easiest emotion to recognize, with *fear* being the most difficult.

5 ANALYZING IMPLICIT CUES

Having noted that females are quicker and better at recognizing (especially negative) emotions, we hypothesized that these behavioral differences should also manifest via implicit cues in the form of brain responses and eye movements during emotional face processing. This section describes the features extracted and methods employed to examine these cues.
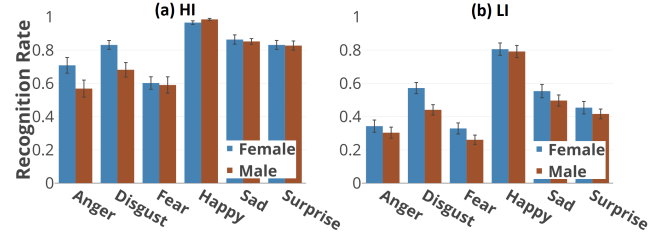
5.1 EEG analysis

We extracted epochs for each *trial* (4.5s of stimulus-plus-fixation cross viewing time @ 128 Hz sampling rate, see Fig. 2), and the 64 pre-stimulus fixation samples were used to remove DC offset. This was followed by (a) EEG band-limiting to within 0.1–45 Hz, (b) removal of noisy epochs via visual inspection, and (c) independent component analysis (ICA) to remove artifacts relating to eye-blinks, and eye and muscle movements. Muscle movement artifacts in EEG are mainly concentrated between 40–100 Hz [5, 13, 34]. While most artifacts were removed upon EEG band-limiting, the remaining were removed manually via inspection of ICA components. Finally, 7168 (14 electrodes × 128 Hz × 4s) dimensional feature matrices were input to the classifier following dimensionality reduction via principal component analysis (PCA) to retain 90% input variance.

5.1.1 ERP analysis. Prior works have observed event-related potential[2] (ERP) based gender differences from EEG data recorded with lab-grade sensors [29, 33, 60]. Specifically, [29] notes enhanced negative ERPs for females in response to negative emotional stimuli. However, capturing ERPs with commercial devices is challenging due to their low signal-to-noise ratio. Fig. 5 presents the P300, visual N100 and N400 ERP components in the occipital O1 and O2 electrodes (see Fig. 1) corresponding to the HI and LI faces. Note that the occipital lobe represents the *visual processing center* in the brain, as it contains the primary visual cortex.

Comparing O1/O2 male and female ERPs for positive (H,Su) and negative (A,D,F,Sa) emotions, no significant difference can be observed between male positive and negative ERP peaks for HI or LI faces (see columns 3,4). However, we observe stronger N100 and P300 peaks in the negative female ERPs for both HI and LI faces (columns 1,2). Also, a stronger female N400 peak can be noted for HI faces consistent with prior findings [29]. Lower male N100 and P300

---

[2]ERP denotes *average* EEG response over multiple trials. P300 (positive) and visual N100, N400 (negative) are examples of ERP components which respectively peak around 300, 100 and 400 ms post stimulus onset.
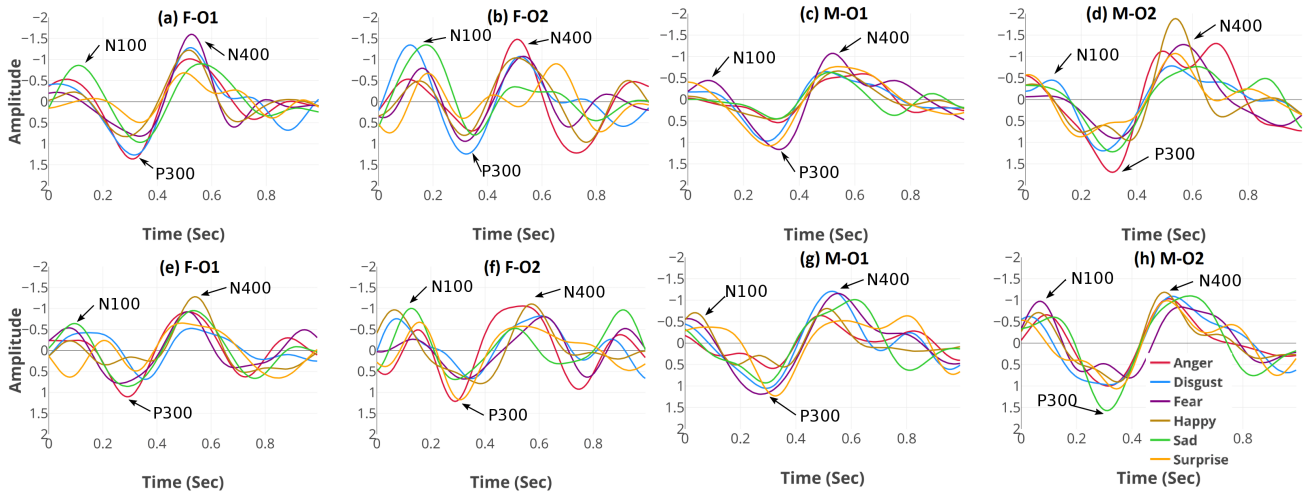
Figure 5: ERPs for HI morphs (a–d) and LI morphs (e–h): (from left to right) O1 and O2 ERPs for females and males.y**-axis shows ERP in microvolts, refer (h) for legend. ERP data is plotted upside down (view in color and under zoom).**

latencies can be observed for positive HI emotions, with the pattern being more obvious at O2. Also, lower male N400 latencies can be noted at O2. The positive vs negative ERP difference for females is narrower for LI faces, which may be attributed to the greater difficulty in identifying LI emotions. This is further corroborated by the fact that LI faces generally produce weaker ERPs at O1 and O2 than HI faces. Overall, the observed ERPs confirm that gender differences in emotional face processing can be effectively captured by the *Emotiv* EEG device.

5.2 Eye-tracking analysis

Gender differences in gaze patterns during emotional face processing have been observed by prior works [42, 49] using high-end eye trackers. In this work, we used the low-cost *Eyetribe* device with 30Hz sampling rate to synchronously record eye movements along with the EEG signals, and compare male and female eye movement patterns. To compute *fixations* from raw gaze positions estimated by the tracker, we used the EyeMMV toolbox [22] and considered the transition from one fixation to another as a *saccade*. Upon extracting fixations and saccades, we adopted the features employed for valence recognition in [40], namely, saccade orientation, top-ten salient coordinates, saliency map and histograms of a) saccade slope, b) saccade length, c) saccade velocity, d) fixation duration, e) fixation count and f) saliency to compute an 825-dimensional feature vector for our analyses.

5.2.1 Fixation analysis. To validate observations on the fixating patterns of males and females, we computed the fixation duration distribution over six facial regions, namely, eyes, nose, mouth, cheeks , forehead and chin. Fig. 6 presents the males and female fixation duration distribution over these facial regions for the various emotions considering (HI+LI) faces. For both men and women, the time spent examining the eyes, nose and mouth account for over 80% of the total fixation duration, with eyes (≈45%) and nose (≈30%) attracting maximum attention as observed in [42]. Focusing on the emotion-wise distributions, a high proportion of eye fixations is
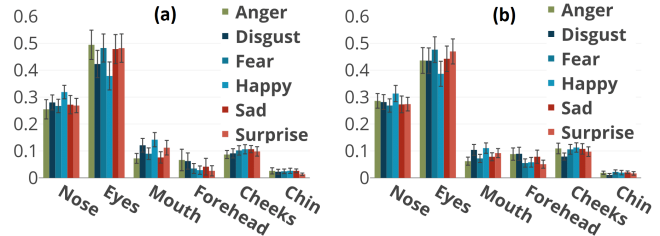

Figure 6: Fixation duration distributions: Region-wise fixation distribution across different emotions for (a) females and (b) males. Error bars denote unit standard error.

noted for all emotions except *happy*, while relatively higher visual attention is noted on the nose and mouth for this emotion consistent with the findings in [42] implying that these regions encode key emotional information concerning happiness. Overall, females were found to examine the eyes and mouth more for emotional cues, with males fixating more on the nose even though none of the differences were significant. The next section presents ER and GR results achieved with EEG and eye-based features.

5.3 Experiments and Results

5.3.1 Experimental settings. We separately examined EEG and eye-based responses to HI and LI morphs for emotion and gender recognition. Specifically, we evaluated ER and GR performance upon feeding (a) only EEG, (b) only eye and (c) concatenated EEG+eye features (termed *early fusion* or EF) to classifiers. Additionally, we probabilistically fused the EEG and eye-based classifier outputs (termed *late fusion* or LF) via the W_{est} procedure outlined in [21], and briefly described as follows. From the EEG and eye-based classifier outputs, the test sample is labeled based on the larger $P_i = \sum_{i=1}^{2} \alpha_i^* t_i p_i$, where i indexes the two modalities, p_i's denote posterior individual classifier probabilities for the 'high' and 'low' classes (we only consider binary GR and ER) and $\{\alpha_i^*\}$ are the optimal weights maximizing test performance, as determined via a

2D grid search. If F_i denotes training performance for modality i, then $t_i = \alpha_i F_i / \sum_{i=1}^{2} \alpha_i F_i$ for given α_i.

Classifiers and training data. For ER and GR, we considered the Naive-Bayes (NB), linear SVM (LSVM) and radial-basis SVM (RSVM) classifiers. NB is a linear generative classifier that assumes class-conditional feature independence to estimate the posterior $p(C/X)$, where C and X denote class-label and feature vector for the test sample. LSVM and RSVM respectively denote the linear and radial basis kernel versions of the popular SVM classifier. For both ER and GR, we only fed (EEG and eye) data corresponding to *correct* trials where users correctly recognized the presented facial emotion.

Performance measure. For benchmarking ER/GR performance, we considered the area under curve (**AUC**) performance measure. The AUC represents area under the ROC curve plotting true and false positive rates, and a random classifier will correspond to an AUC score of 0.5. Also, AUC is an appropriate metric while evaluating classifier performance on data with unbalanced positive and negative class sizes. As we are attempting recognition with few training data, we report ER/GR results over five repetitions of 10-fold cross validation (CV) (*i.e.*, total of 50 runs). CV is typically used to overcome the *overfitting* problem on small datasets, and the optimal SVM parameters were determined from the range $[10^{-4}, 10^{4}]$ via an inner 10-fold CV on the training set.

5.3.2 Emotion Recognition. As explicit behavioral responses (Sec.4) and ERP-based (Sec.5.1.1) analyses suggest that (a) females are especially sensitive to negative facial emotions, and (b) there are cognitive differences in the processing of positive and negative emotions for females and males, we modeled ER as a binary classification problem, with an objective to distinguish between positive and negative emotional classes.

Fig. 7 presents valence recognition (VR) results obtained with male, female and male+female unimodal and multimodal features. Expectedly, better user-centered VR is achieved overall on examining responses to HI emotions (minimum AUC = 0.51 for HI vs 0.49 for LI morphs). Also, best VR performance is achieved with female data (mean AUC = 0.58 with HI morphs) as compared to male (mean AUC = 0.56, HI morphs) and male+female data (mean AUC = 0.543 for HI morphs) considering both unimodal and multimodal features. Interestingly, eye-based features (peak AUC = 0.64 with male data) are more discriminative for VR than EEG features (peak AUC = 0.53 with female data). Evidently, fusing the eye and EEG information is not beneficial even though a simple concatenation of the EEG and eye features is found to be superior to probabilistic fusion of unimodal results.

The obtained results are competitive against (and in some cases better than) prior approaches. Eye based features are found to achieve 52.5% valence recognition accuracy in [40], where emotions are induced by presenting diverse types of emotional scenes to viewers while only emotional faces employed in this work. Also, prior neural studies [3, 33] which employ music and movie content for emotion elicitation achieve only around 60% VR with lab-grade sensors. Our VR results are creditable especially considering that low-cost EEG and eye tracking devices are used in this work and that our objective was not to induce emotions in viewers *per se*– the discriminability instead arises from the manner in which facial
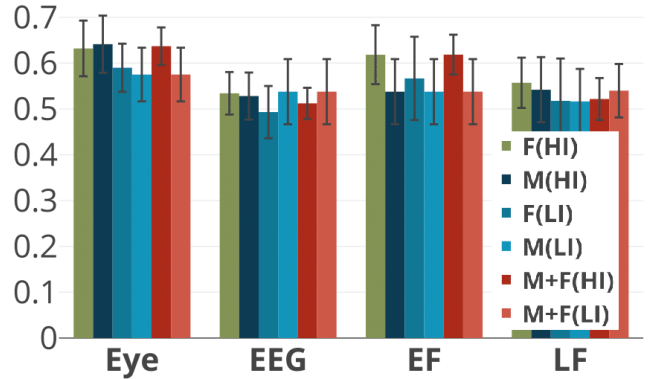

Figure 7: ER Results: valence recognition results with HI and LI morphs.

emotions are *perceived* by viewers. Superior performance of eye-based features may be attributed to the fact that in addition to the difference in fixation duration observable for *happy* (Fig. 6), the saccade and saliency based statistics are able to capture discrepancies in gazing patterns for positive and negative facial emotions [40]; in contrast, EEG data are vectorized and analyzed *as is* without any explicit feature extraction.

5.3.3 Gender Recognition. Table 2 presents eye and EEG-based GR results obtained with HI and LI emotions. Specifically, GR AUC scores with various modalities and classifiers[3] corresponding to *all* and *emotion-specific* trials are tabulated. Clearly better than chance GR is achievable with both EEG and eye-based features. An immediate observation from the tabulated results is that unlike for ER, very similar GR performance is noted for both HI and LI emotions, implying that gender specific differences are also effectively encoded in implicit viewer responses to mildly emotive faces. This is consistent with the explicit user behavioral results in Fig.4, where women are found to outperform men at recognizing LI facial emotions.

Focusing on specifics, EEG features considerably outperform eye-based features for GR and higher AUC scores are obtained with *emotion-specific* (exclusively negative emotion) features as compared to *emotion agnostic* features. This finding is consistent with Fig.4 and prior works such as [29, 60] which have noted significant differences in female neural responses to negative facial emotions. Particularly, the best EEG AUC scores for HI and LI morphs are obtained for the anger and disgust emotions respectively, for which the greatest difference in recognition rates (of 0.14 and 0.13) can be observed from Fig.4. These findings are salient and interesting, as Fig.4 and Table 2 present two different phenomena, namely, gender differences in ER and discriminability of male and female EEG patterns.

Surprisingly, multimodal fusion performs inferior to EEG/eye features. Also, marginally superior GR results are achieved with early fusion, suggesting that perhaps little complementary information is encoded between the EEG and eye-based features. Finally, RSVM is found to produce the best GR performance in most conditions implying that the extracted EEG and eye features are better separable in a higher-dimensional feature space.

---

[3]with which optimal GR results are achieved for a given condition

5.3.4 Spatio-temporal EEG analysis. As EEG represents multi-channel time-series data, we also performed spatial and temporal analyses on the EEG signals to discover (i) the key EEG electrodes that capture gender differences in emotional face processing, and (ii) the critical time window for GR over the 4s of stimulus presentation (or visual emotional processing).

Fig.8(a,b) show the temporal variance in GR performance for HI and LI morphs when the 4s emotional face viewing time is split into four 1s-long windows. These results are computed considering emotion-specific EEG epochs and with information from all 14 electrodes. Overall, no consistent temporal trends concerning gender discriminability can be noted from the plots. However, higher AUC scores over temporal windows W1–W4 are achieved with LI morphs as compared to HI morphs. Also, slightly higher AUC metrics are noted for negative emotions especially with LI morphs, and the best and worst scores are obtained for the *fear* and *happy* emotions. Results of spatial analyses to identify the top three electrodes that best encode gender differences in emotion processing (EEG data over all 4s is used for this analysis.) are shown in Fig.8(c). Again, superior GR is noted with LI morphs in this scenario, and five of the

six electrodes with the best AUC scores correspond to the *frontal lobe*. This result is consistent with the findings in [31], which notes that gender differences in emotion processing are best encoded in the prefrontal cortex.

6 DISCUSSION & CONCLUSION

Of late, there has been considerable interest in the use of technologies such as EEG and eye-tracking to study the cognitive and emotional behavior of users via implicit behavioral cues. Even as potential applications range from affect estimation [3, 20], prediction of erroneous actions in collaborative settings [51] and evaluating the design of data visualization interfaces [38], many of these studies have been limited to lab settings with bulky and intrusive sensors constraining the manner in which users can express themselves.

A key impediment towards employing off-the-shelf sensing devices, which enable naturalistic user behavior in interactive applications has been their poor signal quality, and skepticism regarding the nature of 'useful' information they actually encode. In this regard, we believe that this work adds on to a small body of works such as [51] that have demonstrated the encoding of meaningful information with commercial sensors.

In particular, our work employs the wireless and wearable *Emotiv* EEG device along with the low-cost *Eyetribe* eye tracker to discover gender differences in cognitive and gazing behavior during a facial emotion recognition task. The validity of our study is endorsed by the multiple correlations observed between *explicit* user responses and their *implicit* behavioral patterns. To begin with, women are found to recognize negative facial emotions quickly and more accurately in our experiment, and correspondingly, higher female ERP amplitudes are noted for negative emotions; peak GR performance with EEG, eye features and their fusion is also noted for negative emotions. The maximum difference in female and male recognition rates for HI and LI morphs are noted for the angry and disgust emotions, and the peak GR AUC scores are also noted for these emotions.

Overall, reliable GR with EEG (peak AUC of 0.71), eye features (peak AUC of 0.63) and their fusion (peak AUC of 0.62), as well as valence recognition (peak AUC of 0.64 with eye data) is achieved using the proposed framework. The use of minimally intrusive and low-cost devices for studying user behavior makes our findings ecologically valid, and enhances the possibility of large-scale user profiling. However, limitations in preprocessing like ICA and visual rejection that require a trained eye exist. An alternate to this could be to adopt machine learning methods like deep learning which is now providing end-to-end system architectures with feature extraction incorporated.

The larger aim of this work is to predict user demographics via implicit and anonymous behavioral cues so as to implement intelligent advertising, recommendation systems and mental health monitoring systems(for disorders like Alexithymia).Some of the applications are described here. A recommendation system could be built by evaluating user engagement in advertisements/movies–movie reviews can be highly impacted by certain sequences in it, such reviews are often biased and do not provide a holistic feedback of the entire movie. Using our emotion recognition framework, one can identify the sequences where the audience were distracted,

	AUC	HI	LI
All	EEG(NB)	**0.714 ± 0.002**	**0.600 ± 0.005**
	EYE(NB)	0.493 ± 0.013	0.481 ± 0.017
	Early fusion(RSVM)	0.522 ± 0.035	0.524 ± 0.022
	Late fusion(LSVM)	0.549 ± 0.022	0.523 ± 0.035
Emotion wise — EEG (NB)	A	**0.708 ± 0.064**	0.580 ±0.074
	D	0.673 ± 0.055	**0.696 ± 0.062**
	F	0.643 ± 0.059	0.596 ± 0.089
	H	0.696 ± 0.047	0.668 ± 0.046
	Sa	0.674 ± 0.048	0.634 ± 0.064
	Su	0.692 ± 0.048	0.636 ± 0.071
EYE (NB)	A	**0.601 ± 0.021**	0.565 ± 0.031
	D	0.577 ± 0.011	**0.632 ± 0.009**
	F	0.595 ± 0.015	0.535 ± 0.029
	H	0.560 ± 0.021	0.538 ± 0.017
	Sa	0.539 ± 0.015	0.605 ± 0.030
	Su	0.555 ± 0.008	0.555 ± 0.018
EF (RSVM)	A	0.555 ± 0.021	0.581 ± 0.037
	D	0.535 ± 0.024	**0.622 ± 0.025**
	F	**0.618 ± 0.011**	0.619 ± 0.041
	H	0.575 ± 0.017	0.597 ± 0.009
	Sa	0.540 ± 0.021	0.598 ± 0.024
	Su	0.579 ± 0.013	0.574 ± 0.014
EF (RSVM)	A	0.543 ± 0.062	0.571 ± 0.081
	D	0.542 ± 0.044	**0.597 ± 0.093**
	F	0.519 ± 0.029	0.597 ± 0.170
	H	0.526 ± 0.031	0.508 ±0.017
	Sa	**0.573 ± 0.076**	0.584 ± 0.102
	Su	0.562 ± 0.073	0.568 ± 0.125

Table 2: Table showing Gender Recognition results over different modalities and their fusion for HI and LI morphs using NB-NaiveBayes,RSVM-SVM with RBF kernel and LSVM-SVM with linear kernel.
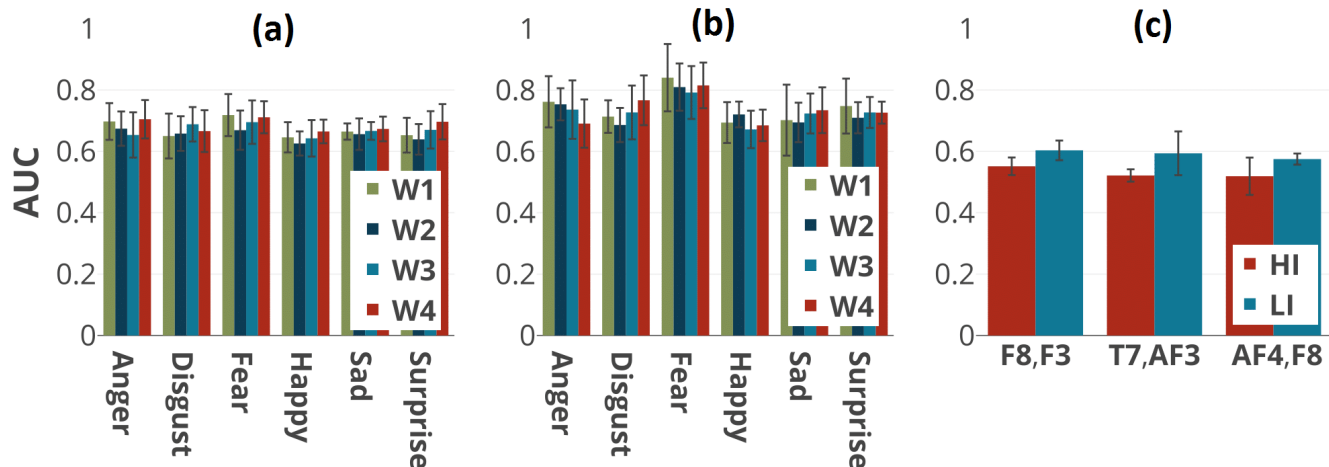
Figure 8: (a,b) Spatio-temporal EEG analyses: GR results for HI and LI morphs over temporal windows (W1–W4). (c) Top 3 electrodes based on AUC metric for HI and LI morphs. Error bars denote unit standard error.

for a better rating system and a feedback system for an immersive cinematic experience. Emotional human faces (especially happy and anger) are ubiquitous on e-commerce websites specifically in categories like clothing and accessories. These can act as stimuli to obtain EEG response and predict the gender to recommend appropriate products. As the window based analysis has shown, the stimulus time could be reduced to close to a second which is practically more feasible system so as to speedup the GR and provide a smooth customer experience. This work represents a proof of concept study, and one can expect more robust and accurate results with larger training data and sophisticated machine learning approaches. Furthermore, computers need to understand and adapt to user emotions for effective human-computer interaction, and gender-specific emotional aspects can be effectively utilized by affective interfaces for manipulating user behavior.

ACKNOWLEDGEMENT

REFERENCES

[1] 2013. Biometric Security Poses Huge Privacy Risks. (2013). https://www.scientificamerican.com/article/biometric-security-poses-huge-privacy-risks/
[2] 2016. How eye tracking gives players a new experience in video games. (2016). http://theconversation.com/how-eye-tracking-gives-players-a-new-experience-in-video-games-54595
[3] M Abadi, R Subramanian, SM Kia, P Avesani, I Patras, and N Sebe. 2015. DE-CAF: MEG-based Multimodal Database for Decoding Affective Physiological Responses. *IEEE Transactions on Affective Computing* 6, 3 (2015), 209–222.
[4] Hillel Aviezer, Ran R Hassin, Jennifer Ryan, Cheryl Grady, Josh Susskind, Adam Anderson, Morris Moscovitch, and Shlomo Bentin. 2008. Angry, Disgusted, or Afraid? *Psychological Science* 19, 7 (2008), 724–732.
[5] Shaibal Barua and Shahina Begum. 2014. A Review on Machine Learning Algorithms in Handling EEG Artifacts. In *The Swedish AI Society (SAIS) Workshop.*
[6] John N Bassili. 1979. Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *Journal of personality and social psychology* 37, 11 (1979), 2049–2058.
[7] Maneesh Bilalpur, Seyed Mostafa Kia, Tat-Seng Chua, and Ramanathan Subramanian. 2017. Discovering Gender Differences in Facial Emotion Recognition via Implicit Behavioral Cues. In *Affective Computing & Intelligent Interaction.*
[8] D. H. Brainard. 1997. The Psychophysics Toolbox. *Spatial Vision* 10 (1997), 443–446.

[9] Patrizio Campisi and Daria La Rocca. 2014. Brain Waves for Automatic Biometric-based User Recognition. *IEEE Transactions on Information Forensics and Security* 9, 5 (2014), 782–800.
[10] Mary Czerwinski, Desney S. Tan, and George G. Robertson. 2002. Women Take a Wider View. In *ACM Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 195–202.
[11] Edwin Dalmaijer. 2014. *Is the low-cost EyeTribe eye tracker any good for research?* Technical Report. PeerJ PrePrints.
[12] Nancy L Etcoff and John J Magee. 1992. Categorical perception of facial expressions. *Cognition* 44, 3 (1992), 227–240.
[13] Theo Gasser, Jan C. Schuller, and Ursula Schreiter Gasser. 2005. Correction of muscle artefacts in the EEG power spectrum. *Clinical Neurophysiology* 116, 9 (2005), 2044 – 2050.
[14] Judith A Hall and David Matsumoto. 2004. Gender differences in judgments of multiple emotions from facial expressions. *Emotion* 4, 2 (2004), 201–206.
[15] Jessica K Hall, Sam B Hutton, and Michael J Morgan. 2010. Sex differences in scanning faces: does attention to the eyes explain female superiority in facial expression recognition? *Cognition and Emotion* 24, 4 (2010), 629–637.
[16] Alan Hanjalic and Li-Quan Xu. 2005. Affective Video Content Representation. *IEEE Trans. Multimedia* 7, 1 (2005), 143–154.
[17] Bruce D. Homer, Elizabeth O. Hayward, Jonathan Frye, and Jan L. Plass. 2012. Gender and Player Characteristics in Video Game Play of Preadolescents. *Computers in Human Behavior* 28, 5 (2012), 1782–1789.
[18] Hideo Joho, Jacopo Staiano, Nicu Sebe, and Joemon M Jose. 2011. Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents. *Multimedia Tools and Applications* 51, 2 (2011), 505–523.
[19] Harish Katti, Ramanathan Subramanian, Mohan Kankanhalli, Nicu Sebe, Tat-Seng Chua, and Kalpathi R Ramakrishnan. 2010. Making computers look the way we look: exploiting visual attention for image understanding. In *ACM Int'l conference on Multimedia.* 667–670.
[20] Sander Koelstra, Christian Mühl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdan, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31.
[21] Sander Koelstra and Ioannis Patras. 2013. Fusion of facial expressions and EEG for implicit affective tagging. *Image and Vision Computing* 31, 2 (2013), 164–174.
[22] Vassilios Krassanakis, Vassiliki Filippakopoulou, and Byron Nakos. 2014. Eye-MMV toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. *Journal of Eye Movement Research* 7, 1 (2014).
[23] Peter J. Lang, Margaret M. Bradley, and B. N. Cuthbert. 2008. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Technical Report A-8. The Center for Research in Psychophysiology, University of Florida, Gainesville, FL.
[24] Oliver Langner, Ron Dotsch, Gijsbert Bijlstra, Daniel H. J. Wigboldus, Skyler T. Hawk, and Ad van Knippenberg. 2010. Presentation and validation of the Radboud Faces Database. *Cognition and Emotion* 24, 8 (2010), 1377–1388.
[25] Chul Min Lee and Shrikanth S Narayanan. 2005. Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing* 13, 2 (2005), 293–303.
[26] Ming Li, Kyu J. Han, and Shrikanth Narayanan. 2013. Automatic Speaker Age and Gender Recognition Using Acoustic and Prosodic Level Information Fusion.

Computer Speech and Language 27, 1 (2013), 151–167.

[27] Xiang Li, Peng Zhang, Dawei Song, Guangliang Yu, Yuexian Hou, and Bin Hu. 2015. Eeg based emotion identification using unsupervised deep feature learning. (2015).

[28] Wei Lun Lim, Olga Sourina, and Lipo Wang. 2015. MIND - An EEG Neurofeedback Multitasking Game. IEEE Computer Society, 169–172.

[29] C. Lithari, C. A. Frantzidis, C. Papadelis, Ana B. Vivas, M. A. Klados, C. Kourtidou-Papadeli, C. Pappas, A. A. Ioannides, and P. D. Bamidis. 2010. Are Females More Responsive to Emotional Stimuli? A Neurophysiological Study Across Arousal and Valence Dimensions. *Brain Topography* 23, 1 (2010), 27–40.

[30] Wei Liu, Wei-Long Zheng, and Bao-Liang Lu. 2016. Multimodal Emotion Recognition Using Multimodal Deep Learning. *arXiv preprint arXiv:1602.08225* (2016).

[31] Erin B McClure, Christopher S Monk, Eric E Nelson, Eric Zarahn, Ellen Leibenluft, Robert M Bilder, Dennis S Charney, Monique Ernst, and Daniel S Pine. 2004. A developmental examination of gender differences in brain engagement during evaluation of threat. *Biological psychiatry* 55 (2004 Jun 1 2004), 1047–55.

[32] Barbara Montagne, Roy P C Kessels, Elisa Frigerio, Edward H F de Haan, and David I Perrett. 2005. Sex differences in the perception of affective facial expressions: do men really lack emotional sensitivity? *Cognitive Processing* 6, 2 (2005), 136–141.

[33] Christian Muhl, Brendan Allison, Anton Nijholt, and Guillaume Chanel. 2014. A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Computer Interfaces* 1, 2 (2014), 66–84.

[34] Suresh Muthukumaraswamy. 2013. High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. *Frontiers in Human Neuroscience* 7 (2013), 138.

[35] Choon Boon Ng, Yong Haur Tay, and Bok-Min Goi. 2012. Vision-based Human Gender Recognition: A Survey. *CoRR* abs/1204.1611 (2012).

[36] Phuoc Nguyen, Dat Tran, Xu Huang, and Wanli Ma. 2013. Age and gender classification using EEG paralinguistic features. In *IEEE EMBS Conference on Neural Engineering*. 1295–1298.

[37] David Passig and Haya Levin. 2001. The Interaction between Gender, Age, and Multimedia Interface Design. *Education and Information Technologies* 6, 4 (2001), 241–250.

[38] Evan M M Peck, Beste F Yuksel, Alvitta Ottley, Robert JK Jacob, and Remco Chang. 2013. Using fNIRS brain sensing to evaluate information visualization interfaces. In *ACM Conference on Human Factors in Computing Systems*. 473–482.

[39] Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA, USA.

[40] Hamed R.-Tavakoli, Adham Atyabi, Antti Rantanen, Seppo J. Laukka, Samia Nefti-Meziani, and Janne Heikkila. 2015. Predicting the Valence of a Scene from Observers' Eye Movements. *PLoS ONE* 10, 9 (2015), 1–19.

[41] Qazi Rahman, Glenn D Wilson, and Sharon Abrahams. 2004. Sex, sexual orientation, and identification of positive and negative facial affect. *Brain and Cognition* 54, 3 (2004), 179–185.

[42] M W Schurgin, J Nelson, S Iida, H Ohira, J Y Chiao, and S L Franconeri. 2014. Eye movements during emotion recognition in faces. *Journal of Vision* 14, 13 (2014), 1–16.

[43] Jeremy D. Schwark, Igor Dolgov, Daniel Hor, and William Graves. 2013. Gender and Personality Trait Measures Impact Degree of Affect Change in a Hedonic Computing Paradigm. *International Journal of Human Computer Interaction* 29, 5 (2013), 327–337.

[44] Abhinav Shukla, Shruti Shriya Gullapuram, Harish Katti, Karthik Yadati, Mohan Kankanhalli, and Ramanathan Subramanian. 2017. Affect Recognition in Ads with Application to Computational Advertising.

[45] Marie L Smith, Garrison W Cottrell, Frederic Gosselin, and Philippe G Schyns. 2005. Transmitting and decoding facial expressions. *Psychological Science* 16, 3 (2005), 184–189.

[46] Ramanathan Subramanian, Divya Shankar, Nicu Sebe, and David Melcher. 2014. Emotion modulates eye movement patterns and subsequent memory for the gist and details of movie scenes. *Journal of vision* 14, 3 (2014), 1–18.

[47] Ramanathan Subramanian, Julia Wache, Mojtaba Abadi, Radu Vieriu, Stefan Winkler, and Nicu Sebe. 2016. ASCERTAIN: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing* (2016).

[48] Ramanathan Subramanian, Victoria Yanulevskaya, and Nicu Sebe. 2011. Can computers learn from humans to see better?: Inferring scene semantics from viewers' eye movements. In *ACM Int'l Conference on Multimedia*. 33–42.

[49] Susan Sullivan, Anna Campbell, Sam B Hutton, and Ted Ruffman. 2015. What's good for the goose is not good for the gander: age and gender differences in scanning emotion faces. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* (2015), 1–6.

[50] M. van Vliet, A. Robben, N. Chumerin, N. V. Manyakov, A. Combaz, and M. M. Van Hulle. 2012. Designing a brain-computer interface controlled video-game using consumer grade EEG hardware. In *Biosignals and Robotics for Better and Safer Living (BRC)*. 1–6.

[51] Chi Thanh Vi, Izdihar Jamil, David Coyle, and Sriram Subramanian. 2014. Error Related Negativity in Observing Interactive Tasks. In *ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3787–3796.

[52] Gail J Walker-Smith, Alastair G Gale, and John M Findlay. 1977. Eye movement strategies involved in face perception. *Perception* 6 (1977), 313–326.

[53] Hee Lin Wang and Loong-Fah Cheong. 2006. Affective understanding in film. *IEEE Trans. Circ. Syst. V. Tech.* 16, 6 (2006), 689–704.

[54] Susan Wiedenbeck, Valentina Grigoreanu, Laura Beckwith, and Margaret Burnett. 2006. Gender HCI: What About the Software? *Computer* 39 (2006), 97–101.

[55] Yingxiao Wu, Yan Zhuang, Xi Long, Feng Lin, and Wenyao Xu. 2015. Human Gender Classification: A Review. *arXiv preprint arXiv:1507.05122* (2015).

[56] S. Yang and F. Deravi. 2012. On the Effectiveness of EEG Signals as a Source of Biometric Information. In *Int'l Conference on Emerging Security Technologies*. 49–52.

[57] Wenhao Zhang, Melvyn L. Smith, Lyndon N. Smith, and Abdul R. Farooq. 2016. Gender and gaze gesture recognition for human-computer interaction. *Computer Vision and Image Understanding* 149 (2016), 32–50.

[58] Wei-Long Zheng, Bo-Nan Dong, and Bao-Liang Lu. 2014. Multimodal emotion recognition using EEG and eye tracking data. In *IEEE Int'l Conference on Engineering in Medicine and Biology Society*. 5040–5043.

[59] Wei-Long Zheng, Jia-Yi Zhu, Yong Peng, and Bao-Liang Lu. 2014. EEG-based emotion classification using deep belief networks. In *Int'l Conference on Multimedia and Expo*. IEEE, 1–6.

[60] Marzia Del Zotto and Alan J Pegna. 2015. Processing of masked and unmasked emotional faces under different attentional conditions: an electrophysiological investigation. *Frontiers in psychology* 6 (2015), 1691.