A System for Non-intrusive Affective Assessment in the Circumplex Model from Pupil Diameter and Facial Expression Monitoring

Sudarat Tangnimitchok[1([envelope])], Nonnarit O-larnnithipong[1], Neeranut Ratchatanantakit[1], Armando Barreto[1], Francisco R. Ortega[2], and Naphtali D. Rishe[2]

[1] Electrical and Computer Engineering Department, Florida International University, Miami, FL 33199, USA
{stang018,nolar002,nratc001,barretota}@fiu.edu
[2] School of Computer and Information Sciences, Florida International University, Miami, FL 33199, USA
fortega@CS.edu, ndr@acm.org

Abstract. This paper outlines a system for non-intrusive estimation of a user's affective state in the Circumplex Model from monitoring the user's pupil diameter and facial expression, obtained from an EyeTech TM3 eye gaze tracker (EGT) and a RGB-D camera (KINECT) respectively. According to previous studies, the pupillary response can be used to recognize "sympathetic activation" and simultaneous "parasympathetic deactivation", which correspond to affective arousal. Additionally, tracking the user's facial muscle movements as he or she displays characteristic facial gestures yields indicators to estimate the affective valence. We propose to combine both types of information to map the affective state of the user to a region on the Circumplex Model. This paper outlines our initial implementation of such combined system.

Keywords: Affective computing · Facial expression recognition
RGB-D camera · Eye-Gaze tracking

1 Introduction

It has been two decades since Affective Computing pioneer Rosalind Picard envisioned a new generation of computers that could interact with their human users at an affective level [5]. However, the complete fulfillment of that goal remains elusive and, already well into the XXI Century, the everyday use of affective computing remains limited. The difficulties associated with the actual implementation of an affective computing system might be best appreciated if one considers the 3 fundamental tasks that must be performed to fully animate the performance of an affective computing system (affective computer), as outlined by Hudlicka [7]: These tasks can be described as:

1. Affect Sensing and Recognition
2. User Affect Modeling/Machine Affect Modeling
3. Machine Affect Expression

The affective sensing and recognition tasks aim at making the machine aware of the affective state of the human user. This will require sensing some observable manifestations of that affective status and recognizing (or "cataloging") the state, so that, then, the machine may determine (by following some pre-programmed interplay guidelines) which affective state it should adopt in response, and, further, the type of affective expression that it should present to the user. Those initial stages of the process, however, may involve some of the major challenges that are presented for the implementation of a fully-functional affective computing system. In fact, Picard identified "Sensing and recognizing emotion" as one the key challenges that must be conquered to bring the full promise of affective computing concepts to fruition [6] (Fig. 1).
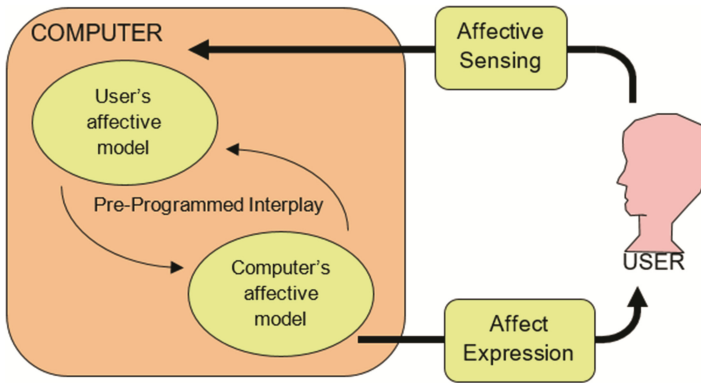
Fig. 1. Simplified diagram showing the interaction between the key processes in affective computing identified in [7]. (Diagram reproduced from [8])

In the pursuit of solutions for that important challenge, there have been may approaches proposed. Specifically, a wide variety of mechanisms have been suggested for affective sensing. Some research groups have attempted the assessment of user affective states using streams of data that are commonly available in contemporary computing systems, such as video from the user's face, audio from the user's voice and text typed by the user on the keyboard.

Zeng et al. [25], provided an interesting survey of relevant systems that use video and/or audio, to estimate the user's affective state. Most vision-driven approaches are based in the known changes that occur in the geometrical features (shapes of eye, mouth, etc.) [10] or facial appearance features (wrinkles, bulges, etc.) [11] of the subject, according to different affective states. Cowie et al. associated acoustic elements to prototypical emotions [9]. Some other groups explored the coordinated exploitation of audio-visual clues for affective sensing [12]. Liu et al. focused on the utilization of text typed by the user for affective assessment [13]. Approaches in this area of work include "Keyword Spotting" (e.g., [14]); "Lexical Affinity" (e.g., [15]); "Statistical Natural Language Processing" (e.g., [16]); etc.

Other groups have attempted to identify the physiological modifications that are directly associated with the affective states and transitions in human beings, and have

proposed methods for sensing those physiological changes in ways that are non-invasive and unobtrusive to a computer user. The reconfiguration experimented by a human subject as a reaction to psychological stimuli is controlled by the Autonomic Nervous System (ANS), which innervates many organs and structures all over the body. The ANS can promote a state of restoration in the organism, or, if necessary, cause it to leave such a state, favoring physiologic modifications that are useful in responding to the external demands.

The Autonomic Nervous System coordinates the cardiovascular, respiratory, digestive, urinary and reproductive functions according to the interaction between a human being and his/her environment, without instructions or interference from the conscious mind [17]. According to its structure and functionality, the ANS is studied as composed of two divisions: The Sympathetic Division and the Parasympathetic Division. The Parasympathetic Division stimulates visceral activity and promotes a state of "rest and repose" in the organism, conserving energy and fostering sedentary "housekeeping" activities, such as digestion [17]. In contrast, the Sympathetic Division prepares the body for heightened levels of somatic activity that may be necessary to implement a reaction to stimuli that disrupt the "rest and repose" of the organism. When fully activated, this division produces a "flight or fight" response, which readies the body for a crisis that may require sudden, intense physical activity. An increase in sympathetic activity generally stimulates tissue metabolism, increases alertness, and, from a global point of view, helps the body transform into a new status, which will be better able to cope with a state of crisis. Parts of that re-design or transformation may become apparent to the subject and may be associated with measurable changes in physiological variables. Variations in sympathetic and parasympathetic activation produce physiological changes that can be monitored through corresponding variables, providing, in principle, a way to assess the affective shifts and states experienced by the subject. Parasympathetic and sympathetic activations have effects that involve numerous organs or subsystems, appearing with a subtle character in each of them.

Therefore, one approach to affective sensing might be based on monitoring the changes in observable variables that are brought about by an imbalance in the sympathetic-parasympathetic equilibrium introduced by sympathetic activation. These changes can then be matched to the fundamental types of states for which each of these divisions of the Autonomic Nervous System prepare us (The sympathetic response prepares us for "fight or flight", whereas the parasympathetic response sets us up for "rest and response"). Accordingly, the predominance of sympathetic activity can very well be taken as an indicator of "arousal", represented on the vertical axis of Russell's Circumplex Model of Affect [3]. It is, indeed, common to experience acceleration of our heart rate (evidence of sympathetic activation) both, while we take a crucial test *and* when our favorite sports team is winning a match (Fig. 2).
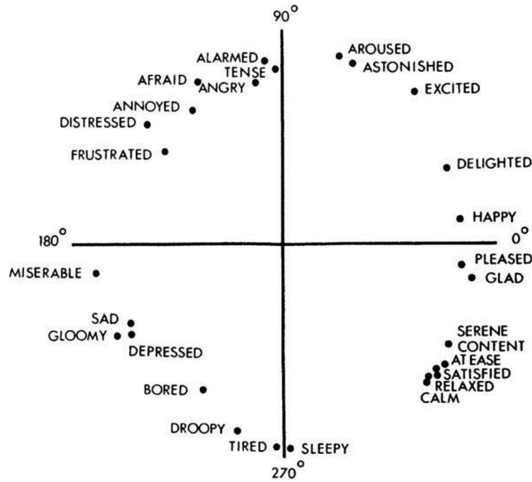
**Fig. 2.** A Circumplex Model of Affect (taken from [3])

Much of our previous work has focused on signal processing methods to estimate a level of sympathetic activation using data recorded from non-invasive physiological sensors, such as Electro-Dermal Activity (EDA), also referred to as "Galvanic Skin Response" (GSR), and, most promising due to its complete unobtrusiveness, Pupil Diameter (PD) monitoring, using infrared video analysis (commonly used in eye gaze tracking, EGT equipment).

However, a more helpful characterization of the user's affective state would also require an indication of the "valence" (horizontal axis in the Circumplex Model). This paper outlines the current direction we have taken to integrate a completely unobtrusive affective assessment system that supplements the arousal estimation provided by pupil diameter monitoring with valence indications derived from the monitoring and classification of key facial features, made possible by the video and depth sensors working in synergy within the KINECT sensor. In the following sections, the paper describes: The rationale and implementation of our arousal assessment through pupil diameter monitoring; The mechanisms used to obtain valence indications from the measurements performed by the KINECT module; and the way in which we are integrating both these modules. The last sections of the paper include some concluding remarks and reflections on the way ahead in the development of this research.

2   Arousal Assessment by Pupil Diameter

As indicated above, our approach to assessing the level of arousal experienced by the subject is through the monitoring of the pupil diameter, measured, in real time, by many eye gaze trackers (EGTs). This approach, in fact, targets the estimation of "sympathetic activation" (and simultaneous parasympathetic deactivation) in the Autonomic Nervous System (ANS). Formerly, our group has explored the monitoring of pupil diameter from a computer user, utilizing an ASL-504 eye gaze tracker, which reports the estimated

pupil diameter in pixels (integer values), for the assessment of affective states in the user [18]. This approach has a strong anatomical and physiological rationale. The diameter of this circular aperture is under the control of the ANS through two sets of muscles. The sympathetic ANS division, mediated by posterior hypothalamic nuclei, produces enlargement of the pupil by direct stimulation of the radial dilator muscles, which causes them to contract [19]. On the other hand, pupil size decrease is caused by excitation of the circular pupillary constriction muscles innervated by the parasympathetic fibers. The motor nucleus for these muscles is the Edinger-Westphal nucleus located in the midbrain. Sympathetic activation brings about pupillary dilation via two mechanisms: (i) an active component arising from activation of radial pupillary dilator muscles along sympathetic fibers and (ii) a passive component involving inhibition of the Edinger-Westphal nucleus [20]. Our rationale is also supported by other independent experiments in which pupil diameter has been found to increase in response to stressor stimuli. Partala and Surakka used sounds from the International Affective Digitized Sounds (IADS) collection [21] to provide auditory affective stimulation to 30 subjects, and found that the pupil size variations corresponded to affectively charged sounds [22].
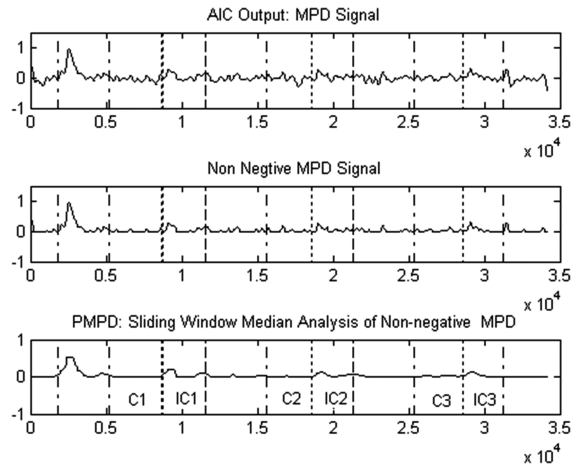
Fig. 3. (From [24]) The bottom panel shows the increased in the Processed Modified Pupil Diameter (PMPD) signal, which correspond to application of stressor ("Incongruent Stroop") stimuli, IC1, IC2 and IC3.

In our current work, we are obtaining measurements of the pupil diameters from both the left and the right eyes at a rate of 30 measurements per second, using a desktop infrared eyegaze tracker, the Eyetech TM3. This eyegaze tracking device operates (in part) by isolating the area of the pupil from the images captured by an infrared camera. The demarcation of the pupil edge is possible because the aperture of the pupil appears as a particularly dark region in the infrared images captured by the infrared camera ("Dark Pupil operation"). It is from that demarcated pupil geometry that the pupil diameter is estimated, in real time. Further details of the "Dark Pupil" principle of operation for eye gaze trackers can be found in the article by Morimoto and Mimica [23].

In our previous work [24], we have verified that an enlargement of the pupil diameter is observed when the subject experiences sympathetic activation from exposure to stressor stimuli ("incongruent" Stroop word presentations), therefore providing further support for the rationale of the combined system described in this paper. Figure 3 shows some of the results obtained.

In this figure, the elevations in the processed signal (PMPD), other than the initial transient at the beginning of the record, are seen to correspond with the intervals labeled "IC1", "IC2" and "IC3", which were the intervals of the experiment when the subject was presented with "incongruent" Stroop word presentations. The details of the experiment, as well as the method used to minimize the impact of potential pupil variations due to illumination changes, can be found in [24].

3   Valence Estimation from Analysis of Facial Features Through KINECT

Humans rely heavily on visual perception for affective sensing; especially when recognizing facial expressions. In general, we recognize an object in front of us by comparing its shape and features with those of objects we learned in the past. Similarly, recognizing human facial expressions can be achieved through the observation of prototypical changes in facial muscles. For example, we may recognize that a person is "happy" because he or she is "smiling".

To supplement our proposed arousal estimation through pupil diameter, and define a 2-dimensional location in the Circumplex Model of Affect [3], we propose a way to estimate the valence (horizontal axis of the model) by using facial expression as the indicator to determine a person's pleasure or displeasure state.

The Facial Action Coding System (FACS) [2], provides a strong foundation for facial gesture recognition. By deconstructing the anatomic components of a facial expression into the specific Action Units (AU), it is possible to code the facial expressions of known affective significance on the basis of the contraction and relaxation of facial muscles. These associations can be leveraged in recognizing affective states from facial gestures. Humans do this through their intrinsic visual perception. For example, we may infer that a person is "happy" by observing the way the corners of his/her mouth are lifted, or the shape of his/her eyes becomes narrower when a person smiles.

In this study, we utilize a Kinect V2 device, which includes a high resolution RGB-D camera, to extract features from a detected face image using its provided APIs. As part of its software framework, its Face APIs enables a wide variety of functionalities, including the delivery of 94 unique "shape units" to create meshes [1] that fit and track a human face in real-time.

It also provides facial points marking important locations such as eyes, cheek, mouth, etc. This allows the tracking of the movement of facial muscles in a way similar to the placement of physical markers on the user face, but in a less intrusive way. The analyzed face feature results then are continuously being updated in the programming object called "FaceFrameResult", as listed in Table 1 [4]. From this list, we are focusing on the features: "Happy", "Engaged", and "LookingAway". Our main interest is in the

feature "Happy" as an indication of the pleasure or displeasure of the subject, while the other two features tell us if a user is engaging the system or not.

Table 1.   List of face features

Heading level	Example
Happy	Happiness data is present
RightEyeClosed	Right eye data is present
LeftEyeClosed	Left eye data is present
MouthOpen	Mouth open data is present
MouthMoved	Mouth movement data is present
LookingAway	Looking away data is present
Glasses	Glasses data is present
FaceEngagement	Face engagement data is present

There are three possible output values for each feature in Table 1, which are "Yes", "No", and "Maybe". We interpret the values of the "Happy" feature as follows: We interpret "Yes" as positive (pleasure), "Maybe" as neutral and "No" as negative (displeasure), hence, obtaining a basic estimation of valence. The next sections provide further details on the combined implementation of our arousal and valence estimation approaches. They will also describe how the results from both subsystems are mapped to the coordination in the Circumplex Model of Affect.

4   Implementation

In this study, we use two devices (hosted by two different computers) to obtain pupil diameter and facial expression.). Both computers communicate through an Ethernet link, using the TCP/IP protocol to share data between them. The system is shown in Fig. 4.

4.1   Pupil Diameter Acquisition

The TM3 eye gaze tracker, from EyeTech Digital Systems is used to obtain pupil diameter data. Its operation requires two steps of initialization, prior to its actual use. First, we run a test program to view the camera stream and fine-tune the angle and location of the TM3 so that it captures an adequate image of both eyes of the user. Secondly, we run a calibration program, where sixteen targets will be shown on the screen one by one. The user will be asked to maintain their head position and direct his/her gaze to the current target until the next one is shown. The process is repeated umtil all 16 targets are gazed upon. After the calibration is done, a calibration file will be generated and saved for the later use (See Fig. 5). Finally, we run a program called Gazeinfo2 to collect our eye gaze information. After the "listen" on-screen button (see Fig. 6) is clicked the program will act like a server waiting for another computer to send a request and then respond back with pupil diameter data.

Fig. 4. Entire system including Kinect V2 (on top of the screen) and TM3 (in front of the computer).
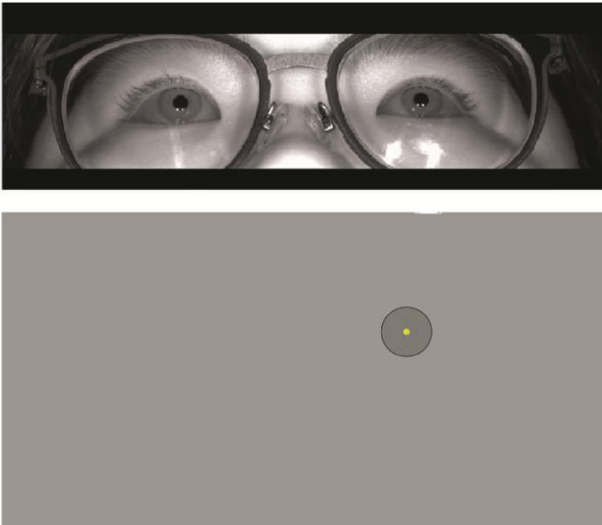
Fig. 5. Two steps of preparation: Adjust position of TM3 (top), Calibration process (bottom)
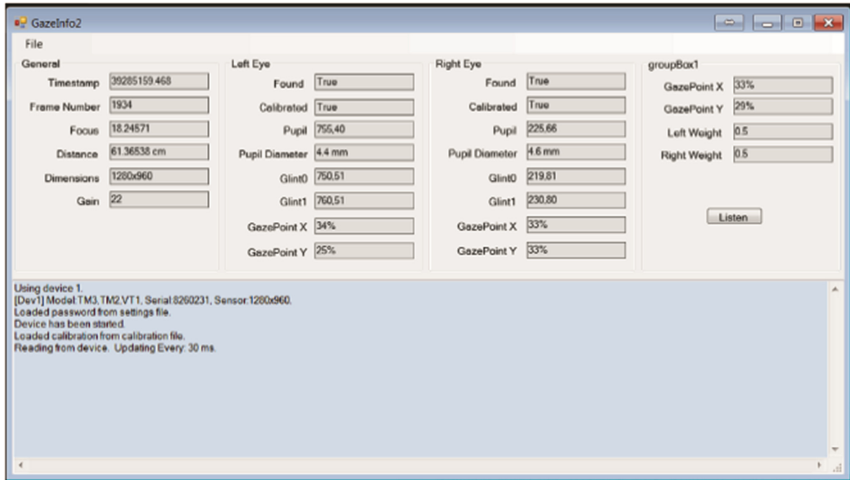
Fig. 6. Graphic interface of the Gazeinfo2 program

4.2 Facial Expression Acquisition

As we already mentioned how Kinect V2 estimates Facial expressions previously, this section describes our own program called "HD_Face", built on the Window Presentation Foundation (WPF) framework (See Fig. 7), which interacts with Kinect V2. Once Kinect V2 detects a user, a violet marker will appear on top of user's face in the video screen. This indicates that Kinect V2 is now collecting the user's facial expression. On the bottom right of the window, the facial expression indicators will now flash in red when
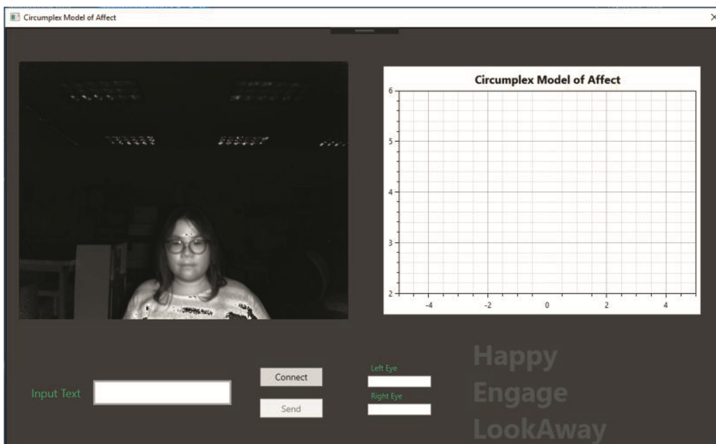

Fig. 7. The user interface of HD_Face program is shown. The top left panel displays the video from the infrared camera. The top right shows a plot of Circumplex Model of Affect. The bottom left is where the communication section is located. Lastly, on the bottom right is where the pupil diameter fetched from another computer and the Facial expression indicator are displayed.

they are asserted by Kinect V2 (For example, "Happy" will flash if the user smiles, as shown in Fig. 8). The other two facial expression indicators work in the same way. (Fig. 9). These two additional indicators provide information that help qualify the validity of the "Happy" indicator. For example, if the system knows that the user is "LookingAway", the absence of a smile detection should not directly be mapped to negative valence.
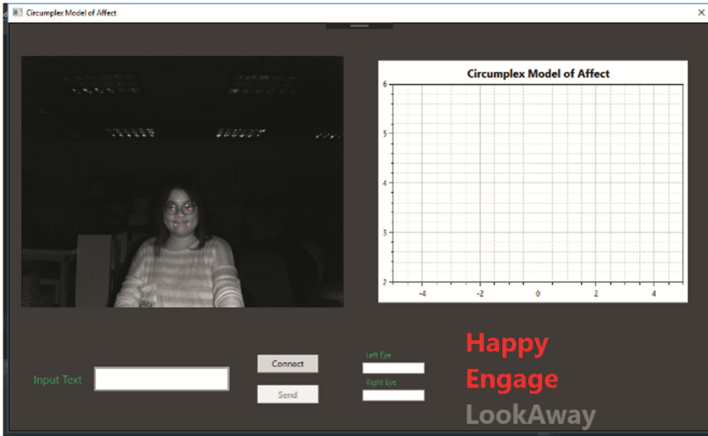
Fig. 8. Example of the facial expression indicator "Happy" flashing in red when a user displays a happy expression.
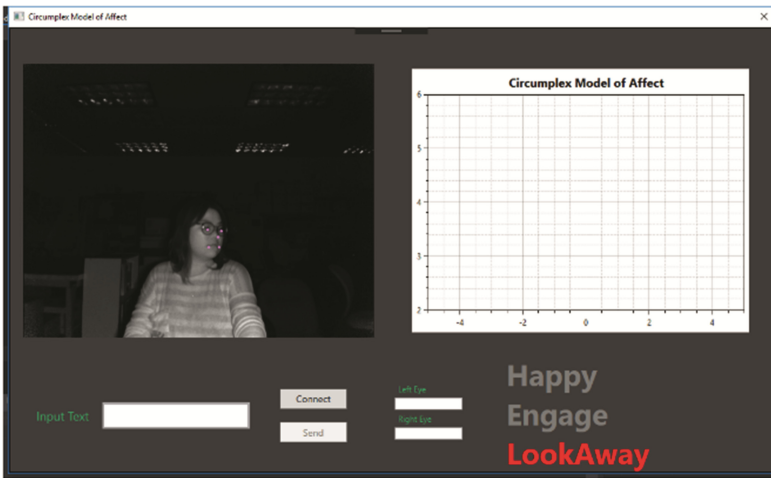
Fig. 9. Example of the facial expression indicator "LookAway" flashing in red when a user looks away.

4.3   Plotting a Circumplex Model of Affect

After making sure that the TM3 Eye Gaze Tracker subsystem is running properly and also verifying that the Kinect V2 is detecting the subject's facial expression (violet marker appearing on the face image), the "Connect" Button on the HD_Face is clicked to request the TM3 subsystem to start sending pupil diameter data. After the connection is established, 1-second averages of the pupil diameter values from both left and right eyes will be displayed on the textboxes located to the left of facial expression indicator
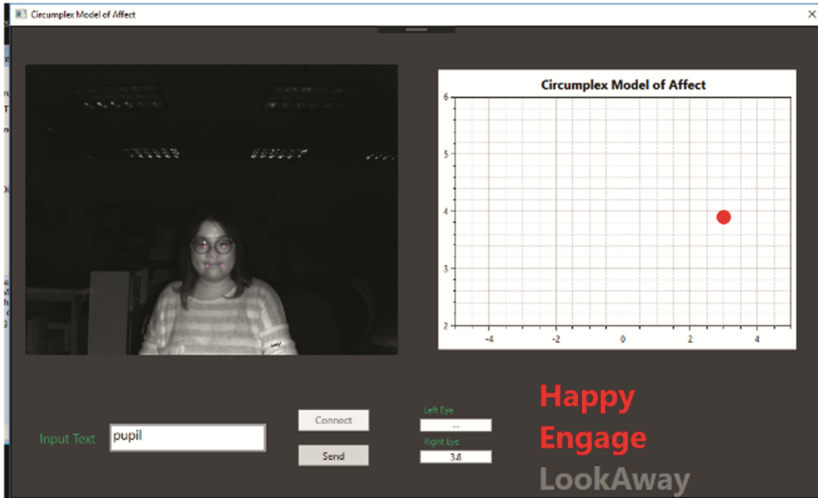

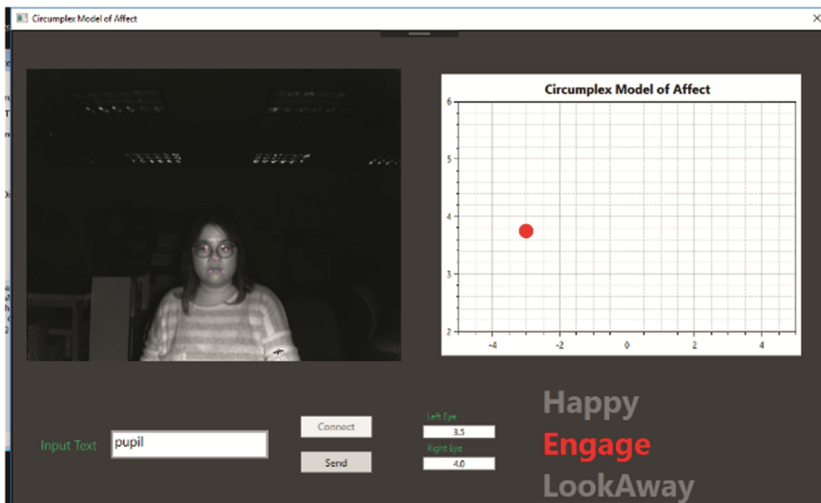**Fig. 10.** Example of plot of the Circumplex Model of Affect when a positive valence was detected

Fig. 11. Example of plot of the Circumplex Model of Affect when a negative valence was detected

section. Using the average pupil diameter values (left + right/2) as the arousal (vertical) coordinate and the scaled "Happy" feature value (Yes = + 3; Maybe = 0; No = −3) as the valence (horizontal) coordinate, a red dot will be continuously plotted in the Circumplex Model of Affect window of the HD_Face screen. The +/−3 scaling value for the "Happy" feature was chosen to satisfy graphical constraints. The pupil diameter is expressed in mm. (See Figs. 10 and 11).

5   Conclusion and Future Work

This paper has outlined our approach to affective state estimation utilizing noninvasive sensors to assess the level of arousal and valence of the affective state of a computer user. Accordingly, these assessments, which can be obtained in real time, can be mapped to a specific region in the Circumplex Model of Affect.

Future aims include the increase of the resolution at which the valence is being assessed, perhaps by performing more specific classification of the facial gestures of the user. Similarly, it will be desirable to define a standard re-scaling procedure of the arousal assessment from pupil diameter values, so that positive and negative values can be assigned to the arousal coordinate in a standardized form.

More robust estimations of the arousal level may be obtained by performing further filtering of the pupil diameter measurements, and by the application of compensatory techniques, such as adaptive noise cancelling, to minimize the undesired impact that variations of environmental illumination may have on the pupil diameter readings.

References

1. Ahlberg, J.: CANDIDE - a parameterized face, 24 May 2012. http://www.icg.isy.liu.se/candide/. Accessed 25 Apr 2017
2. Ekman, P., Freisen, W.V., Ancoli, S.: Facial signs of emotional experience. J. Pers. Soc. Psychol. **39**(6), 1125–1134 (1980). https://doi.org/10.1037/h0077722. Elissa, K.: "Title of paper if known," unpublished
3. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**, 1161–1178 (1980)
4. Microsoft 2017: Face Tracking. https://msdn.microsoft.com/en-us/library/dn782034.aspx. Accessed 25 Apr (2017)
5. Picard, R.: Affective Computing. MIT Press, Cambridge (1997)
6. Picard, R.: Affective computing: challenges. Int. J. Hum.-Comput. Stud. **59**, 55–64 (2003)
7. Hudlicka, E.: To feel or not to feel: the role of affect in human-computer interaction. Int. J. Hum.-Comput. Stud. **59**(1–2), 1–32 (2003)
8. Barreto, A.: Non-intrusive physiological monitoring for affective sensing of computer users. In: Asai, K. (ed.) Human-Computer Interaction New Developments, 1st edn., chap. 4, pp. 85–100. I-Tech, Vienna, August 2008. ISBN 978-953-7619-14-5

9. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.: Emotion recognition in human-computer interaction. IEEE Sig. Process. Mag. **18**(1), 32–80 (2001)
10. Chang, Y., Hu, C., Feris, R., Turk, M.: Manifold based analysis of facial expression. J. Image Vis. Comput. **24**(6), 605–614 (2006)
11. Guo, G., Dyer, C.: Learning from examples in the small sample case - face expression recognition. IEEE Trans. Syst. Man Cybern. Part B **35**(3), 477–488 (2005)
12. Fragopanagos, N., Taylor, J.: Emotion recognition in human-computer interaction. Neural Netw. **18**(4), 389–405 (2005)
13. Liu, H.; Lieberman, H., Selker, T.: A model of textual affect sensing using real-world knowledge. In: Proceedings of the 8th International Conference on Intelligent User Interfaces, Miami, Florida, USA, pp. 125–132. ACM (2003)
14. Elliott, C.: The affective reasoner: a process model of emotions in a multi-agent system. Doctoral Dissertation, Northwestern University (1992)
15. Valitutti, A., Strapparava, C., Stock, O.: Developing affective lexical resources. PsychNol. J. **2**(1), 61–83 (2004)
16. Goertzel, B., Silverman, K., Hartley, C., Bugaj, S., Ross, M.: The baby webmind project. In: Proceedings of the AISB 2000 Symposium, The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (2000)
17. Martini, F.H., Ober, W.C., Garrison, C.W., Welch, K., Hutchings, R.T.: Fundamentals of Anatomy & Physiology, 5th edn. Prentice-Hall, Upper Saddle River (2001)
18. Barreto, A., Zhai, J., Rishe, N., Gao, Y.: Measurement of pupil diameter variations as a physiological indicator of the affective state in a computer user. Biomed. Sci. Instrum. **43**, 146–151 (2007)
19. Steinhauer, S.R., Siegle, G.J., Condray, R., Pless, M.: Sympathetic and parasympa-thetic innervation of pupillary dilation during sustained processing. Int. J. Psychophysiol. **52**, 77–86 (2004)
20. Bressloff, P.C., Wood, C.V.: Spontaneous oscillations in a nonlinear delayed-feedback shunting model of the pupillary light reflex. Phys. Rev. E **58**, 3597–3605 (1998)
21. Bradley, M.M., Lang, P.J.: International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings. Technical report B-2, University of Florida, The Center for Research in Psychophysiology, FL (1999)
22. Partala, T., Surakka, V.: Pupil size variation as an indication of affective processing. Int. J. Hum.-Comput. Stud. **59**, 185–198 (2003)
23. Morimoto, C.H., Mimica, M.R.M.: Eye gaze tracking techniques for interactive applications. Comput. Vis. Image Underst. **98**, 4–24 (2005)
24. Gao, Y., Barreto, A., Adjouadi, M.: Detection of sympathetic activation through measurement and adaptive processing of the pupil diameter for affective assessment of computer users. Am. J. Biomed. Sci. **1**(4), 283–294 (2009)
25. Zeng, Z.; Pantic, M; Roisman, G., Huang, T.: A survey of affect recognition methods: audio, visual and spontaneous expressions. In: Proceedings of the 9th International Conference on Multimodal Interfaces, Nagoya, Aichi, Japan, pp. 126–133. ACM (2007)