Ambiance Signal Processing: A Study on Collaborative Affective Computing

Kaveh Bakhtiyari *†, Mona Taghavi ‡, Milad Taghavi Ж, Jamal Bentahar ‡
Email: academic@bakhtiyari.com contact@monataghavi.com taghavi.milad@gmail.com bentahar@ciise.concordia.ca
* Interactive Systems, Faculty of Engineering, University of Duisburg-Essen, Duisburg, Germany
† Faculty of Engineering & Built Environment, The National University of Malaysia (UKM), Bangi, Malaysia
‡ Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada
Ж Department of Computer Engineering - Software, Islamic Azad University, Parand, Iran

Abstract — **Computational feature recognition is an essential component for intelligent systems to sense the objects and environments. This paper proposes a novel conceptual model, named Ambiance Signal Processing (AmSiP), to identify objects' features when they are not directly accessible by sensors. AmSiP analyzes the surrounding and ambiance of objects/subjects collaboratively to recognize the object's features instead of concentrating on each individual and accessible object. To validate the proposed model, this study runs an experiment with 50 participants, whose emotional state variations are estimated by measuring the surroundings features and the emotions of other people in the same environment. The results of a t-Test on the data collected from this experiment showed that users' emotions were being changed in a course of time during the experiment; however, AmSiP could estimate subjects' emotions reliably according to the environmental characteristics and similar patterns. To evaluate the reliability and efficiency of this model, a collaborative affective computing system was implemented using keyboard keystroke dynamics and mouse interactions of the users whose emotions were affected by different types of music. In comparison with other conventional techniques (explicit access), the prediction was reliable. Although the developed model sacrifices a minor accuracy, it earns the superiority of uncovering blind knowledge about the subjects out of the sensors' direct access.**

Keywords—**Ambient Intelligence; Human Emotion Recognition; Affective Computing; Human-Computer Interaction; Feature Recognition**

I. INTRODUCTION

The objects around us make the environment, and their features provide the ambiance. The objects have a mutual influence on each other since they contain energy and transmit their energy to other objects nearby. The human being is also considered as an object in the environment. Consequently, other objects impact human characteristics including emotions, actions, thoughts, and decisions. A human being is exploring the objects and their characteristics to make a pleasant environment and to enhance their own lifestyle [1]. For example, businesses make the best ambiance as possible for their employees to improve their performance, or another example is designing efficient websites and online advertisements, which resonate with the users effectively, and finally result in a possible conversion. In general, there is a mutual relation between the objects of the same or similar clusters. Therefore, recognition of objects' features and characteristics can help to estimate the other objects' characteristics, which are in the same cluster or nearly similar clusters [2]. The objects of a cluster may vary in type, but they share some similar features or affects. For instance, all objects in a room can be in the same cluster, even though they are not the same or similar objects. In this example, location is the mutual feature of all objects in a cluster. Not only the mutual features but also the mutual affect of features can be used for clustering the objects. This is especially useful when there is no access to all objects, or it is very time-consuming or expensive to investigate all objects around.

Computational feature recognition is useful in many areas such as Artificial Intelligence. Intelligent systems rely heavily on feature recognition techniques to sense and process the environment (such as users, behaviors, images, audio, etc.) to learn and to make proper decisions. Recognizing the features of an object is an essential step for an intelligent system to understand the object.

For decades, researchers have tried to recognize the object features by using different sensors and employing complex mathematical techniques. Sometimes, there is no direct access to the objects, and the system cannot interact with the object (or the user) directly and explicitly. In the absence of direct access, analysis of the environment and the objects nearby can shade away from the out of reach objects [3].

Previously, the similar estimation for feature recognition was possible based on the statistical models (e.g. k-NN algorithm), but the result was a generalized outcome about a group of objects based on the other objects of the same class with a similar type, similar features. Therefore, extracting a feature of a specific object was not simply possible by statistical models, especially when there was no data space of the objects of the same class.

In this study, Ambiance Signal Processing (AmSiP) is proposed to overcome this problem. AmSiP estimates a specific object's properties by considering and analyzing the other relevant objects (not necessarily similar) nearby in the same environment. The main contribution of this research is proposing a conceptual model to:

- Recognize an object's features (e.g. user's emotion), when there is no direct access to the object,
- Identify the other possible available objects and their features in a place, while there are no means to do so.

Affective Computing (AC) is the study of recognizing, interpreting, processing, and simulating human emotions. We considered an AC system via the analysis of the user interactions as a case study. This system has been validated by t-Test and evaluated by Support Vector Machine (SVM) and Artificial Neural Network (ANN). This model has improved the performance of the system by providing more relevant information to estimate the object's features. AmSiP works based on the analysis of social signals of the social relations [4] in social signal processing (SSP) [5, 6], in which the users share the same location proximity. The achieved results are superior in comparison with the conventional models and methodologies which only concentrate on the objects alone to extract the features. In this case study, AmSiP estimates individual users'

978-1-7281-1431-6/19/$31.00 ©2019 IEEE

emotional arousal according to the environmental sound and the other users at the same location when there may not be direct access available to the user. The system has been developed based on the music and users' interaction with mouse and keyboard in order to analyze the details of the environment and ambiance.

The rest of the paper is organized as follows. In the following section, a background study on affective computing is discussed. In section 3, the conceptual model is introduced. Then, in section 4, it follows with the experimental setup to validate the proposed model. In section 5, the results of the experiment are discussed.

II. Affective Computing

AC is one of the challenging, but yet addressable issues which has been receiving a high amount of attention from the researchers in computer science during recent years. Many computer games, applications, and security issues can be healed and tailed with human emotions to improve Human-Computer Interactions (HCI) [3, 7, 8]. There are many techniques and methods available to recognize the human emotions explicitly such as facial expression recognition [9], audio/speech signal processing [10], NLP, Electroencephalography (EEG) [11], heart rate [12], skin conductance [13], etc.

To the best of our knowledge from the current literature on AC, there is no research about how to recognize an individual user's emotions if there is no means of access (e.g. physical, visual, auditory, etc.) to the user. In other words, the user out of a sensor's access will remain a mystery for AC.

As it has been stated earlier, previously statistical analysis was being used to estimate the features of a group based on the other groups. But these methods cannot be simply used in AC. Because based on statistics, estimation is usually about a group rather than an individual, and the target group should be in the same class. Also, simple statistical models usually fail to estimate the user's emotions and behaviors accurately [14]. Even though the statistical model shares some basics with AmSiP, it simply ignores the features and affects of the other objects of different types and the environments. However, AmSiP proposes a solution to this problem, which will be discussed in the next section.

HCI is a very important discipline in AC. The style of typing (keystroke dynamics), mouse movement patterns and touch interactions can also reflect the user's emotions [15]. Different emotions make different styles and patterns of interaction by users, while the users are interacting with computers [16]. In this paper, implicit AC by user interaction analysis is conducted as a case study to validate the proposed model.

III. Conceptual Model of AmSiP

When data and objects are in the same place together, some similar behaviors and parameters are expected to manifest themselves. As a very simple example, most of the people in a festival are expected to be happy. This phenomenon is also applicable to the data. Data mining techniques, machine learning algorithms, and such similar methods are following the same concept. They learn patterns of data, therefore, they can forecast and make decisions for future similar events. A primary set of data, training set, is used for learning purposes. Then there is a new set of data, a test set, which is independent of values from the training set, but dependent on the behaviors of the training set. Therefore, the test set is expected to show the same reaction, and to produce the output similar to the training set. The dependency of data, which may be collected from different people or objects, is in relation to the environment of data collection. When the whole data is collected from a specific cluster, the data interrelations would be increased.

AC has been represented as an application for AmSiP in this research. Fig. 1 resembles the scenario of the model where several people are exposed to an environment consisting of different objects such as music and lightening. Among these people, only two of them are interacting directly with the computers, three of them have no interactions with the available computers but they are accessible, and one is absolutely out of the sensors' scope. In this case, AmSiP tries to estimate the emotions of those four people who are not interacting with the computer individually including the one who is out of the sensors' access.

AmSiP model is defined by Surroundings and Environments. All surroundings (Sur_N) belongs to the global group of the environment (Env) defined in Eq. 1.

$$\{\exists\, O_1 \ldots . O_n \in Sur_N\}.\{\exists\, Sur_1 \ldots . Sur_n \in Env\} \quad (1)$$

Objects are identified as O, where i and j are the indices of set N objects. Each Sur_N contains objects which are similar in the features (f) (Eq. 2) and in their affects (Eq. 3).

$$f\,(O_i \in Sur_N) \approx f\,(O_j \in Sur_N) \quad (2)$$

$$Affect(O_i \in Sur_N) \approx Affect(O_j \in Sur_N) \quad (3)$$

The objects in a different surrounding group may not share similar features or affects. The average of each feature (f_i) of an object in a surrounding can be denoted as F_i, which is relatively the value for the same feature of the object n. Thus, the feature of an object in a surrounding N can be calculated by Eq. 4.

$$F(Sur_N) = \frac{\sum_{i=1}^{n} \alpha_i \frac{\overrightarrow{O_N}.\overrightarrow{O_i}}{\|\overrightarrow{O_N}\| \times \|\overrightarrow{O_i}\|} + \beta_i \frac{\overrightarrow{Affect_{O_N}}.\overrightarrow{Affect_{O_i}}}{\|\overrightarrow{Affect_{O_N}}\| \times \|\overrightarrow{Affect_{O_i}}\|}}{n} \quad (4)$$

Two parameters of α and β are user-defined gain values between 0 and 1, which are used as a weight for each feature or affect. These values can also be calculated based on other factors (such as appraisal models to calculate β) or through machine learning methods.

At last, $F(O_N)$ can be predicted by using the features and affects of the other objects in the surrounding, as in Eq. 5.

$$F(O_N) = f_n(O_N) + \sum_{j=1}^{m} f_j(O_n \in Sur_N) \quad (5)$$

EXAMPLE. Here is a simple example to understand the above equations. Let us assume that there is a user which its

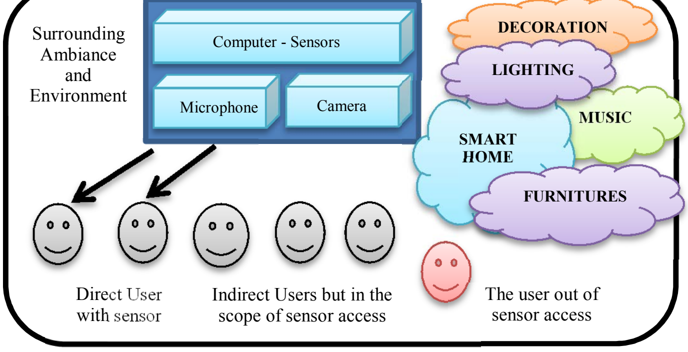


Fig. 1. The proposed conceptual model of AmSiP in an AC system

emotion and gender are unknown to the system. However, the system has access to the user's environment (his/her room) via a camera. In the recorded images, there are two identical objects: red lipstick and a glass of wine on the desk. According to the lipstick presence, it can be estimated that the user might be a female, but the presence of the lipstick does not show any significant affect on the user. Therefore, in this example for the lipstick α (which is for the object's features) is higher than β (which shows the weight of the object's affect). The second object is a glass of wine. This object's features do not represent any useful information about the gender of the user, however, it can have a pleasant affect on the user [17]. Therefore, the *Affect(wine) = pleasure.* In this case, the β value is much higher than α. Selecting the correct values can predict more reliable and accurate results based on how effective the features and affects are.

IV. Experiment Setup

In order to assess the feasibility and evaluate the precision and accuracy of the proposed model in this research, AmSiP estimates individual users' emotional arousal implicitly according to the environmental sound and the other users at the same location when they are not directly accessible to the sensors. Here, we developed a prototype system to process the music played in the environment, and the users' interaction with a mouse and keyboard in order to analyze the details of the environment and ambiance.

AmSiP was tested by building a user profile of 50 users, 25 males and 25 females with an average age of 23 years old, in a 2-hour session. Participants were clustered into two groups of 25 people randomly. They were asked to sit in a computer laboratory and to work with computers regularly for a minimum of 2 hours. They had few options, to read some emotionally neutral textbooks or websites, plus surfing the social networks.

The computers available in the laboratory were equipped with the prototype application developed by C#.NET to record the user's keyboard keystroke dynamics, mouse features (e.g. mouse clicks, mouse movements, mouse sleepiness) and their self-reported emotions after 15 minutes, one hour and two hours. The recorded features consisted of time intervals between the features of interactions.

At the beginning of the test, they verified their neutral emotion. In order to make the environment suitable for the experiment, two sets of music were prepared. Set 1 was a collection of *New Age* music with relaxation theme. Set 2 was a collection of *Jazz* and *Rock* music. Set 1 was played for Group 1, and Set 2 was played for Group 2 for two hours. It is assumed to have users relaxed (low arousal state) with the relaxing music and tensioned/hyper (high arousal state) with the Rock and Jazz music. Thus, the data are expected to be different in each time of recording with the initially generated profile which was created in the first five minutes. After this experiment, user profiling on about 50 users is completed, and it can be used for further analysis.

In the second experiment, 50 participants were selected (80% from the first experiment), and a random set of music was being played to detect and analyze their behaviors and emotions based on the generated profiles in the first experiment using SVM with the Gaussians kernel and ANN. Processing the surrounding objects in a place gives a sense of what is happening in that proximity.

In the following, the methodology of recording the keyboard keystroke dynamics and mouse features are being elaborated.

A. Keystroke Dynamics

Keystroke dynamics are a habitual rhythmic pattern of typing which is usually used for user identification for many years. There are three major features in keystroke dynamics as below [18]:

- Key down-to-down
- Key down-to-up
- Key up-to-down

Besides the above three main features, 15 sub-features can be extracted. The following 15 features are based on digraphs (two-letter combinations) and trigraphs (three-letter combinations) [15]:

- *KF1:* The duration between 1st and 2nd down keys of the digraphs.
- *KF2:* The duration of the 1st key of the digraphs.
- *KF3:* Duration between 1st key up and next key down of the digraphs.
- *KF4:* The duration of the 2nd key of the digraphs.
- *KF5:* The duration of the digraphs from 1st key down to last key up.
- *KF6:* The number of key events that were part of the graph.
- *KF7:* The duration between 1st and 2nd down keys of the trigraphs.
- *KF8:* The duration of the 1st key of the trigraphs.
- *KF9:* Duration between 1st key up and next key down of trigraphs.
- *KF10:* The duration between 2nd and 3rd down keys of the trigraphs.
- *KF11:* The duration of the 2nd key of the trigraphs.
- *KF12:* Duration between 2nd key up and next key down of trigraphs.
- *KF13:* The duration of the third key of the trigraphs.
- *KF14:* The duration of the trigraphs from 1st key down to last key up.
- *KF15:* The number of key events that were part of the graph.

Classification of these features with SVM and ANN classifiers have proved to be effective and promising with the accuracy of 85.2% for nervousness and 91.24% for the fright emotion [15].

B. Mouse Interactions

There is a large amount of research in mouse interaction analysis. They are based on the mouse movements, mouse clicks, and even mouse sleepiness. A mouse movement generates a racing line path, called curve, which may or may not be linear. At every moment, mouse position can be read by its coordinates of (x,y). The combination of coordinates and time can compute the velocity, acceleration, direction, and path. The calculated mouse movement features are listed below [19]:

- The length of the mouse racing line
- The zero crossings
- The maximum deviation of the values
- Average of the racing line values
- The standard deviation of the racing line
- The variance of the racing line
- Correction function of the curve

Application of the above features alone recently obtained an impressive accuracy in emotion recognition between 78.7% for fright and 93% for neutral emotions [18, 20].

V. EVALUATION

The PANAS psychological scale is a standard method for measuring the valence [21]. We have used a similar scale for the arousal dimension in the scale of -2 to 2. The value of 0 is considered as the neutral state of the user. Studies show that higher arousal causes slightly more and faster keystroke dynamics and mouse interactions. On the other hand, relaxed users (low arousal) might have fewer keystroke dynamics and slower mouse movements with less velocity [15, 18]. In order to evaluate the model, the value of 0 is considered as the mean value for the Neutral emotion. The values of -1 and 1 denote Low Arousal (Calming emotions) and High Arousal (Arousing emotions) respectively. Value ranges of [-2,-1) and (1,2] belong to more relaxed (LA-) and aroused users (HA+).

The average of recorded data in the first stage (first five minutes) was considered as a Neutral state. Fig. 2 and Fig. 3 depict users' arousal levels for Group 1 with relaxing music, and Fig. 4 and Fig. 5 show the change of high arousals for Group 2.

A. Measuring Typo-Errors

There is some research to correlate user typo mistakes with their emotional states and arousal levels. It shows that higher arousal increases typo errors (it increases the number of backspace keystrokes), however, lower arousal (calming emotions) causes fewer typo errors [16, 22]. In addition to the recorded arousal levels, typo errors were also measured and averaged for each group. Fig. 6 presents the typo errors occurred by two groups in two hours while typing on the keyboard (if necessary) in the experiment. The typo errors are measured by counting the number of "*backspace*" key pressed.

In Fig. 6, the comparison of typo error rates shows that it would be affected by the emotional changes. As the arousal level increases, the rate of type errors would also be increased. On the other hand, a calming environment causes fewer typo errors in user interactions. The recorded typo errors with timestamps show that the number of errors is being stabilized by the time, and it would not increase or decrease significantly.

B. Investigation of Dependency through Paired t-Test

The *Paired t-Test* is applied to measure the difference significance of the collected data presented in Fig. 2 and Fig. 4. Since the collected data are dependent and was captured from the same persons periodically, the Paired t-Test has been selected to compare the changes in time. The t-Test has been running for 3 times on Calming Ambience and Arousing Ambience. The first test compares the neutral state of the users (level value at 0.0) with the initially collected data. The second test compares the initial data with the data collected in the 1st hour. And the third test compares the collected data at the 1st hour with the 2nd hour. Table I shows the results of the Paired t-Test on the presented data in Fig. 2 and Fig. 4.

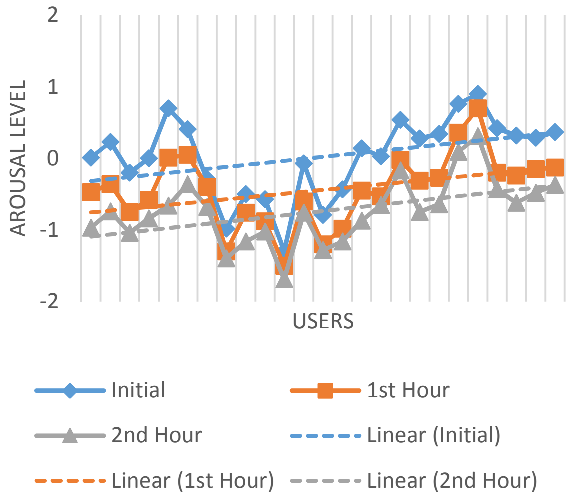


Fig. 2. HV levels of Group 1 with relaxation music

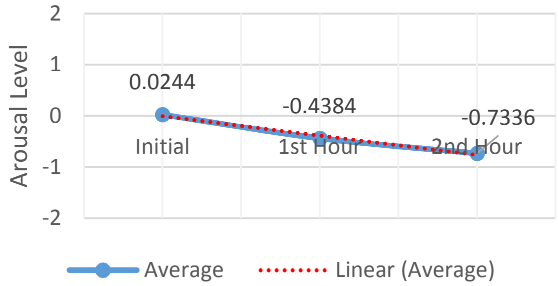


Fig. 3. Average of LA levels change in Group 1 with relaxation music

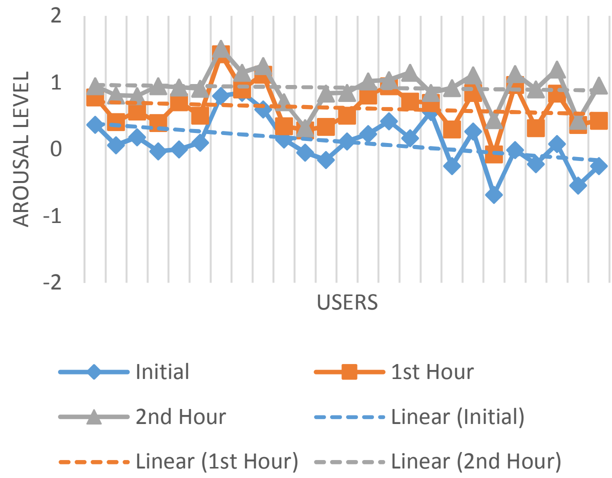


Fig. 4. HV levels of Group 2 with Rock and Jazz music

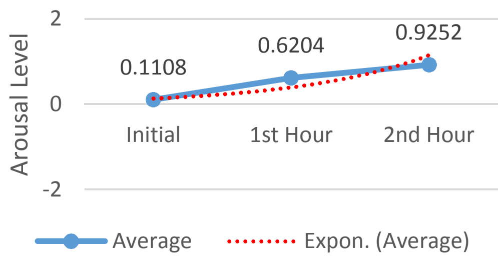


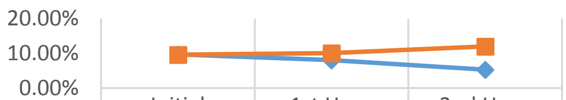
Fig. 5. Average of HV change in Group 2 with Rock and Jazz music

In Table I, the 1st t-Test result shows that there is no significant difference between the Neutral state and the Initial state. This behavior could also be expected as the initial state was recorded a few minutes after the neutral state. Even though the environment has affected the users somehow, but the differences are not statistically significant. The comparison between 2nd and 3rd t-Test results clearly states the higher difference in the 2nd Test. It means that the environment affects the users mostly in the 1st hour. The changes made in the 2nd hour are not as significant as in the 1st hour.

C. *Model Accuracy*

Learning from the change of the user's tension levels during the 2-hour session could provide a predictive model to be used in an AC system as a reference for further predictions. We used SVM classifier with the ANN kernel to evaluate the prediction accuracy of the proposed AmSiP model.

At the very first stage, the system needs to create a base model for each individual user to find out their standard neutral emotional state in the first five minutes. Then the model was created through the users' interactions and changes in their tension levels. Table II provides the accuracies of detection and their false positive (FP) rates based on the AmSiP model. It is presented with the Mean Absolute Error (MAE) threshold of 0.1 (2.5%) for the true positive (TP) values.



	Initial	1st Hour	2nd Hour
Group 1	9.72%	8.04%	5.31%
Group 2	9.63%	10.04%	11.97%

Fig. 6. Typo errors recorded by two groups in different stages

TABLE I. Paired t-Test among various states for low and high arousal

	Neutral State - Initial State	**Initial State - 1st Hour**	**1st Hour - 2nd Hour**
	Low Arousal (LA)		
t-Test	0.2219	14.9458	10.3763
P	0.8262	< 0.0001	< 0.0001
d	0.1100	0.031	0.028
95% CI	(-0.251,0.202)	(0.398,0.526)	(0.236,0.353)
	High Arousal (HA)		
t-Test	1.5071	11.8721	8.5651
P	0.1448	< 0.0001	< 0.0001
d	0.074	0.043	0.036
95% CI	(-0.262,0.040)	(-0.598,-0.421)	(-0.378,-0.231)

TABLE II. Accuracy report with 0.1/4 (2.5%) of MAE threshold for True Positives

	After 1 Hour						
	Relaxation Music				Rock – Jazz Music		
	Detected	Actual	Accuracy		Detected	Actual	Accuracy
LA	38	47	80.85%	**HA**	36	46	76.59%
Neutral	1	2	50%	**LA**	1	2	50%
LA-	0	1	0%	**Neutral**	2	2	100%
FP	11	-	22%	**FP**	11	-	22%
	After 2 Hours						
	Relaxation Music				Rock – Jazz Music		
	Detected	Actual	Accuracy		Detected	Actual	Accuracy
LA	45	48	93.75%	**HA**	43	47	91.49%
Neutral	0	0	Null	**Neutral**	1	1	100%
LA-	1	2	50%	**HA+**	1	2	50%
FP	4	-	8%	**FP**	5	-	10%

AmSiP model could estimate the user's emotion collaboratively by measuring the emotion of other people in the same cluster as well as the affect of other objects. Usually, an easy and simple clustering can be based on a mutual location. The results of t-Test in TABLE I already showed that the users in the same environment share a common emotional pattern. In addition, cross-user emotion detection was performed to compare the detection result, with the explicitly direct (concentrated) emotional arousal recognition. The recorded dataset in two stages was selected for cross-user emotion detection (collaborative emotion estimation). Each emotion of the user was estimated by considering the other users and objects in the same environment. The calculated implicit estimation was quite reliable but with 15% – 20% less accuracy (and higher MAE) than the explicit (concentrated) methodologies introduced in [15, 18, 23].

VI. Discussion

All the classic object processing methods, including image processing, voice recognition, and signal processing concentrate on the objects only. When the object is out of the system access, it would stop further processing because there is no more data to be processed. According to the reported results, the AmSiP model provides a slightly lower accuracy without any knowledge about the object itself, however, the conventional models cannot estimate any values with blind knowledge about the object. The object's features and characteristics may be estimated by analyzing the features and/or affects of other accessible objects by the system. This estimation may not be very accurate, but the system is still able to capture more information about the target which is out of the sensors' access.

Deploying AC systems can be foreseen in the future applications. Through understanding human emotions, computers can offer enhanced interaction with their users by providing suitable services and interfaces [24]. These interactions make more natural communication experience with computers, which results in higher user trust and usability.

AmSiP is a new model of processing the objects to retrieve the information. Exercising AmSiP in AC can bring up a new research topic that might be extended in different areas such as HCI, Recommender Systems, Computational Advertising, etc.

In AmSiP, it is important to identify the most relevant objects, and their features and affects, and to cluster them into various surroundings. However, all identified objects may not be relevant to our target subject. To estimate the value of $f(O_{n+1})$, the objects in the same or relevant group should be considered. Therefore, there would be two major groups of relevant (R_{el}) and irrelevant (I_{rel}) groups of objects (Eq. 6).

$$\exists \begin{Bmatrix} R_{el} = [Sur_p. \dots . Sur_q] \\ I_{rel} = [Sur_{p'}. \dots . Sur_{q'}] \end{Bmatrix} \in Env \quad (6)$$

The features of the relevant objects (R_{el}) are directly required and useful for $f(O_{n+1})$, however, the features of irrelevant objects (I_{rel}) can distract the estimation of $f(O_{n+1})$ and produce noise (Eq. 7). Eliminating the irrelevant objects can enhance the system accuracy.

$$\begin{aligned} &F\ (R_{el})\ \rightarrow f(O_{n+1}) \\ &F\ (I_{rel})\ ! \rightarrow f(O_{n+1}) \end{aligned} \quad (7)$$

Defining the relevant and irrelevant objects for O_{n+1} is itself an open challenge in this model, and it requires further research.

VII. CONCLUSION & FUTURE WORKS

This paper presented a novel conceptual model, which processes and analyzes the relevant actions as well as objects in an environment to extract the other specific object features in the neighborhood. This concept is a step forward from the classic concept with a focused concentration on the accessible object features. To validate the efficiency of AmSiP model in an AC system, several experiments were conducted. The system could validate the users' emotions through their interactions with a computer using keyboard keystroke dynamics and mouse interactions analysis. The t-Test results of this experiment proved that there is a significant influence of the environment on the users. The results of SVM showed that this model is reliable and can detect human emotions in the absence of access to other users. The AC system was tested by forecasting the emotions of a user by considering the other users and possible influences in the same environment. In comparison with other techniques which use direct access methods, the prediction was reliable but with 15% – 20% less accuracy. However, this model has the superiority of identifying the features which the direct techniques are not able to provide.

AmSiP is not a technique, but it is a new concept of looking at the surroundings to obtain the required information implicitly. It shows the importance of the objects and their effects that they were ignored previously in the analysis. This solution works with a concept which says: *"the more objects you have, the more information you may get."* In our future research, we will apply AmSiP on collaborative recommender systems [25] and develop a standard framework for AmSiP.

ACKNOWLEDGEMENT

We would like to thank Prof. Dr.-Ing. Jürgen Ziegler for his review and comments on an early version of this manuscript.

REFERENCES

[1] T. Bosse, R. Duell, Z. Memon, J. Treur, and C. N. van der Wal, "Agent-Based Modeling of Emotion Contagion in Groups," *Cogn Comput,* vol. 7, no. 1, pp. 111-136, 2015.

[2] C. Boldrighini, I. A. Ignatyuk, V. A. Malyshev, and A. Pellegrinotti, "Random walk in dynamic environment with mutual influence," *Stochastic Processes and their Applications,* vol. 41, no. 1, pp. 157-177, 1992.

[3] A. Vinciarelli *et al.*, "Open Challenges in Modelling, Analysis and Synthesis of Human Behaviour in Human–Human and Human–Machine Interactions," *Cogn Comput,* vol. 7, no. 4, pp. 397-413, 2015.

[4] H. H. Kelley and J. W. Thibaut, *Interpersonal Relations: A Theory of Interdependence*. Hoboken, NJ: John Wiley & Sons, 1978, p. 341.

[5] R. A. Calvo, S. D'Mello, J. Gratch, and A. Kappas, *The Oxford Handbook of Affective Computing*. Oxford University Press, 2014.

[6] A. Pentland, "Social signal processing [exploratory DSP]," *IEEE Signal Processing Magazine,* vol. 24, no. 4, pp. 108-111, 2007.

[7] R. Cowie *et al.*, "Emotion recognition in human-computer interaction," *Signal Processing Magazine, IEEE,* vol. 18, no. 1, pp. 32-80, 2001.

[8] J. Röning, J. Holappa, V. Kellokumpu, A. Tikanmäki, and M. Pietikäinen, "Minotaurus: A System for Affective Human–Robot Interaction in Smart Environments," *Cogn Comput,* vol. 6, no. 4, pp. 940-953, 2014.

[9] J. F. Cohn, I. O. Ertugrul, W.-S. Chu, J. M. Girard, L. A. Jeni, and Z. Hammal, "Affective facial computing: Generalizability across domains," in *Multimodal Behavior Analysis in the Wild*: Elsevier, 2019, pp. 407-441.

[10] S. Zhang, S. Zhang, T. Huang, W. Gao, and Q. Tian, "Learning Affective Features With a Hybrid Deep Model for Audio–Visual Emotion Recognition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 28, no. 10, pp. 3030-3043, 2018.

[11] B. Hu, X. Li, S. Sun, and M. Ratcliffe, "Attention Recognition in EEG-Based Affective Learning Research Using CFS+KNN Algorithm," *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* vol. 15, no. 1, pp. 38-45, 2018.

[12] A.-M. Brouwer, E. van Dam, J. B. Van Erp, D. P. Spangler, and J. R. Brooks, "Improving real-life estimates of emotion based on heart rate: a perspective on taking metabolic heart rate into account," *Frontiers in human neuroscience,* vol. 12, 2018.

[13] E. Gatti, E. Calzolari, E. Maggioni, and M. Obrist, "Emotional ratings and skin conductance response to visual, auditory and haptic stimuli," *Scientific Data,* Data Descriptor vol. 5, p. 180120, 06/26/online 2018.

[14] D. Howitt and D. Cramer, *Introduction to Statistics in Psychology*, Fifth Ed. ed. Pearson Education, 2011.

[15] K. Bakhtiyari, M. Taghavi, and H. Husain, "Hybrid affective computing—keyboard, mouse and touch screen: from review to experiment," *Journal of Neural Computing and Applications,* vol. 26, no. 6, pp. 1277-1296, 2015.

[16] F. Monrose and A. D. Rubin, "Keystroke dynamics as a biometric for authentication," *Future Generation Computer Systems,* vol. 16, no. 4, pp. 351-359, 2000.

[17] R. Ferrarini, C. Carbognin, E. M. Casarotti, E. Nicolis, A. Nencini, and A. M. Meneghini, "The emotional response to wine consumption," *Food Quality and Preference,* vol. 21, no. 7, pp. 720-725, 2010.

[18] K. Bakhtiyari, M. Taghavi, and H. Husain, "Implementation of Emotional-Aware Computer Systems Using Typical Input Devices," in *Intelligent Information and Database Systems*, vol. 8397. Bangkok, Thailand: Springer International Publishing, 2014, pp. 364-374.

[19] K. Bakhtiyari and H. Husain, "Fuzzy model of dominance emotions in affective computing," *Journal of Neural Computing and Applications,* vol. 25, no. 6, pp. 1467-1477, 2014.

[20] K. Bakhtiyari and H. Husain, "Fuzzy Model on Human Emotions Recognition," presented at the 12th WSEAS International Conference on Applications of Computer Engineering (ACE), Cambridge, MA, USA, 2013. [Online]. Available: http://arxiv.org/abs/1407.1474.

[21] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *Journal of Personality and Social Psychology,* vol. 54, no. 6, pp. 1063-1070, 1988.

[22] J. S. Lee and D.-H. Shin, "A Study on the Interaction between Human and Smart Devices Based on Emotion Recognition," in *HCI International 2013-Posters' Extended Abstracts*: Springer, 2013, pp. 352-356.

[23] A. Almanji, A. Payne, R. Amor, and C. Davies, "A Nonlinear Model for Mouse Pointing Task Movement Time Analysis Based on Both System and Human Effects," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. PP, no. 99, pp. 1-1, 2014.

[24] R. W. Picard, "Recognizing Stress, Engagement, and Positive Emotion," in *Proceedings of the 20th International Conference on Intelligent User Interfaces*, 2015: ACM, pp. 3-4.

[25] M. Taghavi, J. Bentahar, K. Bakhtiyari, and C. Hanachi, "New Insights Towards Developing Recommender Systems," *The Computer Journal,* vol. 61, no. 3, pp. 319-348, 2018.