Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2014) 000–000

Procedia
Computer Science

www.elsevier.com/locate/procedia

6th International conference on Intelligent Human Computer Interaction, IHCI 2014

Investigating context awareness of Affective Computing systems: A critical approach

Aggeliki Vlachostergiou*, George Caridakis[1], Stefanos Kollias

Image Video and Multimedia Systems Laboratory
National Technical University of Athens
Iroon Polytexneiou 9, 15780 Zografou, Greece

Abstract

Intelligent Human Computer Interaction systems should be affective aware and Affective Computing systems should be context aware. Positioned in the cross-section of the research areas of Interaction Context and Affective Computing current paper investigates if and how context is incorporated in automatic analysis of human affective behavior. Several related aspects are discussed ranging from modeling, acquiring and annotating issues in affectively enhanced corpora to issues related to incorporating context information in a multimodal fusion framework of affective analysis. These aspects are critically discussed in terms of the challenges they comprise while, in a wider framework, future directions of this recently active, yet mainly unexplored, research area are identified. Overall, the paper aims to both document the present status as well as comment on the evolution of the upcoming topic of Context in Affective Computing.

Keywords: Affective Computing; Context Modeling; Context aware systems; Adaptative and intelligent interactive systems; Affect and emotion models; Interaction context representation and annotation

---

1. Introduction

The last 20 years has witnessed a number of important efforts aimed at technologies for modeling, analysis and synthesis of human-human and human-computer interactions (HCI). Due to the fact that human and computer existence have become extremely interwoven, recent developments in that research area have already shifted from HCI to intelligent HCI (iHCI), moving from traditional keyboard and mouse to natural humanlike interactive functions including understanding certain human behaviors such as affective and social signals. In other words, by approaching HCI towards the HCI settings that are based on touch, gesture and movement (e.g. Microsoft Kinect), we have the ability to detect subtleties and changes in the human's communicative behavior and thus to initiate interactions based on this information. Hence, the result is an increasingly growing state-of-the-art in domans such as Affective

* Corresponding author. Tel.:+30-210-772-4052 ; fax: +30-210-772-2492.
[1] George Caridakis is also affiliated with the Department of Cultural Technology and Communication, University of the Aegean.
E-mail address: aggelikivl@image.ece.ntua.gr (A. Vlachostergiou).

Computing (AC), social signal processing, HCI [1], gaming [2], mental health [3], learning technologies [4] etc. However, relatively little attention was paid to the potential impact these technologies can achieve when the contextual aspect is incorporated in human affective behavior analysis systems.

Incorporating context in affective human behavior analysis lies at the intersection of context aware affective computing systems and affective aware intelligent human computer interaction systems since contextual information cannot be discounted in doing automatic analysis of human affective behavior. The contributions of context aware affective computing systems were demostrated during the two recently organized workshops on Context-based Affect Recognition (CBAR 2012 and CBAR 2013 [5]), held in conjuction with SocialCom2012 and ACII 2013 respectively. More specifically, the CBAR 2013 workshop was one of the shortest workshops in ACII 2013 with the most interesting keynotes (Schuller, Gratch) which is indicative of the research area's status. These prominent figures in AC research community provided a valuable source of information discussing about learning context in affect recognition and how context shapes how people interpret the expressions of people and machines. Thus, by tackling the issues of context based affect analysis, i.e. careful study of contextual information and its relevance in context aware systems, its representation, its modeling and incorporation including its effect on the performance of existing affect analysis methods, it is possible to outline a roadmap showing the most important steps still to be made towards real-world affect analysis.

Attempting to formally define context aware affect analysis systems, a starting point would be to investigate how the term context has been defined. "Context" has a multitude of meanings even within the field of Computer Science (CS). To illustrate this, we group different definitions of the term context in the area of artificial intelligence, natural language processing, image recognition, mobile computing, vision-assisted tagging of personal photo albums, object recognition and scene classification, where every discipline has its very own understanding of what context is.

According to the first work which introduced the term context awareness in CS, [6] the important aspects of context are: Who you are with, Where you are, When, What resources are nearby. Thus, context aware systems look at the Who, Where, When and What (the human is doing) entities and use this information to determine Why the situation is occurring. In a similar definition, in [7] authors define context as location, identities of the people around the user, the time of day, season, temperature, etc. Other approaches [8] include context as the user's location, environment, identity and time while others [9,10] have simply provided synonyms for context. For a more extended overview on context-awareness the reader is referred to [11].

In the area of AC, context awareness is recognized as an important element in human-computer interfaces and can be broadly defined as the understanding of the location, the person's identity as well as the type and timing of the HCI [6]. In human affective communication, literature indicates that people evaluate situations based on contextual information such as past visual information [12], general situational understanding, past verbal information [13], cultural background, gender of the participants, knowledge of the general interaction setting in which an emotional phenomenon is taking place [7], discourse and social situations [14,15], to avoid any misinterpretations of the observed affective cues such as facial, vocal or gestural behavior.

As far as real-world, context aware affective computing frameworks is concerned, context is defined as any information that can be used to characterize the situation that is relevant to the interaction between the users and the system [6]. Thus, to incorporate the analysis of our primary means of communicative affective, social and cognitive states, the W4 technology has been extended to the W5+ to fulfill the need for switching from HCI to iHCI. The W5+ formalization is considered as an even more suitable definition, as it summarizes the key aspects of context (Who is involved (e.g. dyadic/triadic interactions among people or one person and a virtual character), What is communicated (e.g., (non)-linguistic message/conversational signal, and emotion), how the information is communicated (the person's affective cues), Why, i.e., in which context the information is passed on, Where the user is, what his current task is, How he/she feels (has his mood been polarized changing from negative to positive?) and which (re)action should be taken to satisfy human's needs, goals and tasks.) [16].

Unfortunately, so far the efforts on human affective behavior understanding are usually context independent due to the fact that the human behavioral signals are easily misinterpreted if the information about the situation in which the shown behavioral cues have been displayed is not taken into account. Thus, up to date the proposed methodology has answered one or more context-related questions such as Who, Where, When, What, Why and How either separately or in groups of two or three using the information extracted from multimodal input streams [17].

The paper differs from the previous overview papers that investigated Affective Computing w.r.t. the contextual aspect by focusing on the latests developments and trends by mostly incorporating a number of representative works introduced after 2009. Thus, this overview attempts to narrow the communicative gap between the highly emotional human and the emotionally intelligent HCI systems that recognize and respond to the affective states of the user by grouping the works exhibited in the subdomains of integrating context on affect production, interpretation and analysis respectively.

The paper is structured as follows: we first focus on the affective representation approaches for context and the corpora featuring affective context and annotation process (Section 2). We then proceed with exploring the problem domain of context awareness in AC by presenting context aware affect analysis systems whose performance has been improved after incorporating the contextual aspect (Section 3). Section 4 explores a number of challenges and unresolved issues when integrating context on affect production, interpretation and recognition. The paper concludes by discussing the future directions and providing some recommendations to advance the field (Section 5).

2. Affective representation, corpora and annotation for context

How to represent emotions and affect is one of the first decisions to be made prior to creating automatic affect analysis systems. We provide a brief discussion on affective representation approaches in terms of their applicability to context. Additionally, methodologies for incorporating contextual information in affective corpora (e.g. how it ought to be represented, what contextual information is relevant (i.e. is it domain specific or not?)) are also investigated.

2.1. Affective representation for context

There is an increased interest in understanding the relative benefits and limitations of alternative theoretical affect models, cognitive processes and their associated representations in terms of their applicability to context. Three dominant theoretical emotion models have been established in AC: categorical, dimensional and appraisal [17]. The categorical approach claims that there exist a small number of emotions, which are basic and recognized universally. The dimensional approach advocates that the affective states are related to one another, where each basic emotion represents a bipolar entity being a part of the same emotional continuum. The proposed polars are arousal (relaxed vs. arousal) and valence (pleasant vs. unpleasant). However, in view of their suitability to context modeling, emphasis is given to the emotional models based on cognitive appraisal, which characterize emotional states in terms of the detailed evaluations of emotions acquisition and especially implicit methods. For an extended overview on modeling affect, the reader is referred to [18].

Recently a research attempt has been witnessed to argue that another set of psychological models, referred to as componential models of emotion, which are based on the appraisal theory, might be more appropriate for developing context aware frameworks [19], however, how to use the appraisal approach for automatic analysis of affect is an open research problem. In the componential models of emotion, various ways of linking automatic emotion analysis and appraisal models of emotion are suggested. This link aims to enable the addition of contextual information into automatic emotion analyzers, and enrich their interpretation capability in terms of a more sensitive and richer representation. Based on their approach the emotion analysis process is divided into two mapping schemes: expressive features to appraisal variables (first layer) and appraisal variables to emotion label (second layer), providing a number of benefits for automatic emotion analysis [19]. Thus, this latter appraisal based model is more related to the W5+ formalization as it consists of a decomposition attempt of the appraisal process into the two above layers.

2.2. Corpora featuring affective context and annotation

Clearly designed and annotated corpora in terms of context validate context modeling process and its incorporation in AC architectures. The annotation process generally reveals that context is indeed very difficult to model in naturalistic interaction as in these interaction settings each definition might reflect different assumptions on annotating the contextual information. Moreover, due to the fact that there is no agreed data acquisition protocol that would be applied to provide improved results, the task of identifying and extracting contextual information in existing affective corpora becomes even harder. Even for a human expert is difficult to define or identify what constitutes context related to emotion. So far, there are no repositories of context-dependent affective data that could assist in shaping our understanding of the problem itself and shed light into main annotation problems in terms of incorporating context. Few exemptions that satisfy some of the above requirements are driven by the interaction between the human and a virtual agent. For example, in the SEMAINE dataset[20] the whole personality traits acquisition process is driven by the context of human-sensitive artificial listener interaction[1] (user is seated, looking at the agent, agent has a specific style and mood, user is not allowed to ask questions, agent asks the questions, etc.).

Thus, the challenges of achieving this goal begin by applying a general framework for naturalistic data collection and annotation that includes a context layer. This domain is still in its early stages and no major efforts have been made so far for the collection of context-dependent affective data specifically aimed at the analysis of contextual information. Most of the works in literature use data originally aimed at different purposes and annotated ad hoc for satisfying the needs of the performed experiment each time. Obtaining ground truth can be very challenging and requires a strict data annotation protocol regarding the global definition of context by the annotators. However, group interactions and decision making involve a large variety of aspects and no standard annotation or data collection protocol seem to be easy to implement. There is currently a significant need for such data in order for the field to be able to move forward.

Such challenges include the dynamic nature of emotions, the ambiguity in categorization and the high subjectivity in emotion perception. We are aware of two recent context's annotation schemes[14,15]. In the former, the authors aim to explore the relation between the user and the social context in terms of the perceived involvement. Focus is given on the behavior of the four users in the group context. Their behavior during the free conversation is analyzed first individually and then as a group, annotating the degrees of their engagement and involvement, using discrete labels, while in the latter work[15], the authors applied a similar approach for annotating changes in involvement, using continuous labels to show any "increase", "decrease" or "no change" respectively during the session. Finally, in another recent work[21], the discussion revolved around how the knowledge of context (the presence of audiovisual channels and the knowledge about the previous interaction) influences the multimodal annotation within a natural HCI.

3. Context aware affect systems

Following context integration on affect production and context incorporation in emotion corpora, this section reviews context awareness in automatic affect analysis systems, both single-modal and multimodal. In[22] a unimodal context aware affect analysis system for short-term context modeling in dyadic interactions is proposed, where context is defined as the speech cues from the past utterance of the speaker, while in[23] a multimodal phoneme recognition system used Bidirectional Long Short Term Memory (BLSTM) networks to incorporate arbitrarily large amount contextual data from past and future contextual information. Overall, in terms of emotion analysis approaches, Bidirectional context-sensitive approaches proved to outperform methods that do not consider context from both past and future observations such as Recurrent Neural Networks (RNNs) and Hidden Markov Models (HMMs). However, such conclusions depend on the design of the corpus and may not cover the entire spectrum of human affectively enhanced communication in terms of interaction modeling.

Moreover, in terms of visual signals, the interpretation of the observed facial expression[24] has been also attempted in the past. Facial expressions are accordingly displayed in a particular context, such as location (outdoor, indoor), situation (driving a car or being treated in a hospital), the undergoing task, other people involved, the personality of the expresser[25,17]. However, to the best of our knowledge, no vision-based model takes into account the context of the application for spontaneous expression analysis. A recent exemption is discussed in the work of[26] where the Transferable Belief Model (TBM) is used to easily add one or more context variables in the model of facial expressions classification for the pain analysis application. In this work, several contextual variables can be defined (the place, the task, the answer to a writing question, etc.) to reduce the set of the expected facial expressions. At the same time only "the place" as a contextual variable is used and takes two values depending on whether the expresser is in the hospital or not in order to identify if the videotaped expression is painful or not. Furthermore, the former work consists of an extension of[27], where a context aware clinical system related to pain analysis is described, where the contextual variables are defined as the place, the task, etc.

As far as context aware affective frameworks w.r.t. the educational technology is concerned, based on the so far reported results to infer a child's interest level when playing an educational game and when detecting his levels of confidence, frustration, excitement and interest in naturalistic school settings [18], in most cases the combination of facial with contextual features achieved substantial improvements over the base line accuracies. A representative example is the combination of facial features with posture patterns and contextual information that has been extended to include a skin conductance sensor and a pressure-sensitive mouse [24]. This system predicted self-reported frustration while children engaged in a problem solving task and yielded an accuracy score of 79%, which is a substantial improvement over the 58.3 % accuracy base line. Unfortunately, due to the fact that the latter study did not report single-channel classification accuracy; hence, it is difficult to assess the specific benefits of considering multiple channels.

4. Challenges

With the growing interest in the context awareness in HCI, a better understanding of the "Why" question in a W4-context-dependent manner when analyzing HCI, would provide opportunities for possible fruitful areas of research broader in the field of AC. Thus, at this section, we aim to turn the spotlight to some of the challenges and opportunities related to the intermediate steps of HCI presented below for future innovation. More specifically, this section focuses on the challenges when integrating context on affect production, when incorporating contextual information in affective corpora (e.g. how it should be represented, what type and amount of contextual information is relevant and sufficient respectively) and when incorporating context in context aware affect systems.

The main criticism of the works [28] that attempted to reply to W5+ context questions is that **the methods are not applicable in real-life situations**: Hence, the focus of the research in the field started to shift to automatic affective analysis when integrating context on affect elicitation (explicitly or implicitly) of spontaneous human behavioral events and lately in data collected in the wild [29].

Context-independent affect production: Another limitation of affect production in terms of affect elicitation w.r.t context information is that the data sets collected include affective expressions that are collected in a context-independent environment. However, the nature of affective expressions is not context-independent. On the contrary, it is highly situation dependent [30] and context is critical because it helps to disambiguate the observed facial, vocal or body behavior. Since affective expressions can convey different meanings in different contexts, training such affect detection systems in context-independent environments is unlikely to produce systems that will adapt to new contexts.

Another issue is related to the use of **data collected in the wild in terms of containing faces in unconstrained conditions** [29]: Both psychologists and engineers tend to acquire their data in laboratories and artificial settings [31,32], to elicit explicitly the specific emotional phenomena they want to observe. However, this is likely to simplify excessively the situation and to improve artificially the performance of the automatic approaches. Aligned with the aforementioned trend, there also has been much work in creating large-scale visual sentiment datasets [33,34,35], as the role of visual content has become even more prominent in social media such as Twitter, where the textual description is limited to very short messages and the tagging is completed without any effort from the users [33]. Typically, such choices of vocabulary are defined according to utility for diversity, availability of training material, while semantic concepts (e.g. objects, locations, activities in visual data) can be easily automatic detected. Recent approaches have also turned towards web portals like Flickr and YouTube as user-generated tags as an alternative to manual labels [36].

Incorporating Context as a dimension: This presents particular challenges, as discussed in [37] and any advancement in that front will advance relevant research in analysis of behavioral data in general. Deciding whether context should be treated as an extra dimension is not clear yet. To date, research community has been situated among categorical, dimensional and appraisal-based representations. Although most AC applications seem to require these major approaches, some have argued that componential approaches [19] might be more appropriate for building affective-aware frameworks. However, identifying the appropriate level of representation for practical AC applications is still an unresolved question.

Fusing context with other modalities: It has been proven that the integration of multiple modalities produces superior results in human behavior analysis when compared to single modal approaches. The analysis of context is no different as one can see in [38,24,39,18].Many of the multimodal systems in the field adopt a number of fusion methods that make use of the correlation between different streams [17], however further work is needed along this front.

Contextual design: At which level in the processing stream does contextual information have a role? In the area of HCI, contextual cues play a crucial role for the interpretation of social attitudes as social parameters such as the situation, roles, relations of the persons involved and user's parameters such as the personality, the person's affective state, the person's choice influence the choice and the interpretation of cues to be shown. Thus, there is the challenge of incorporating context when designing computational models of emotion elicitation and emotion expression, taking into account all the necessary ethical considerations and following systematic guidelines for affective modeling[40]. Taking into consideration these difficulties, the reliability of emotion recognition components is expected to improve.

Evaluating context aware affect aware applications: The characterization of the performance of a model is usually based on the reliability of a coding scheme, or measurement instrument, kappa scores (agreement after correcting for chance) which in naturalistic contexts range from poor to fair. To date, there have been no commonly agreed upon protocols for evaluation, nor do benchmark scenarios for testing such technologies appropriately. This is partially due to the fact that systems' requirements might differ in terms of users' preference elicitation effort and therefore users will more likely adopt the system that demands less effort. Thus, it is of crucial importance to conduct comparative user studies of the existing designs among the different models to reveal factors that influence users' perception and attitudes towards a particular system design. Such a clarification will simplify our decision on the emotion model that is needed based on the given context, its development and evaluation.

5. Discussion and Future Directions

Attempting to recognize the communicative intention including affective and cognitive states of the user, the questions "Why" and "How" have been added to shift from the W4 to the W5+ formalization, taking advantage of the multimedia data and social media resources. For example, in social media such as Twitter, Flickr, YouTube and other web portals, a natural progression of context-related questions about the bombing story of 11th of September in U.S. revolves initially around the four out of five "W's" - Who, What, When and Where of the W5+ formalization. In other words, answering the questions: What happened? When did it happen? Where did it happen? Who was involved? and Who said what? defines the context of the incident and is necessary to describe the full story. The question of Why it happened? follows the W4 context formalization and is related to subjective identification and sentiment analysis. Consequently, by capturing the principle quartet of W's and using each as a pillar towards a foundation platform, we are able to tackle the issues of context awareness of Human-Computer analysis to progress towards real-world affect analysis.

Thus,our context-aware framework overview is presented below: The new context incorporation architecture into affect-aware systems detects a set of (visual) semantic concepts from the social media feeds that refer to a physical presence of objects or scenes that define context, focusing on the "Where" concept question. These visual concepts can be used to fill the affective gap and automatically infer the sentiments reflected in an image, providing answers to the "Why" context-related question which is naturally subjective. These lexical concepts can be used by the publicly available online knowledge sources (OKS) in natural language processing such as the General Inquirer[41], the WordNet[42] and the ConceptNet[43] that contain information about words, concepts, or phrases, as well as connections among them. More specifically, the General Inquirer consists of lexical concepts, where each one denotes the presence of a specific property in the term. WordNet organizes lexical concepts in terms of synonymy, meronymy and antonymy, while the latter linguistic knowledge database provides a semantic network of 26 different relations that encode the meaning of the connection between them. Thus, the proposed approach at this point is to take advantage of the verbs and nouns which are closest to affect related words (as determined by the General Inquirer) via these OKS. Unknown words to General Inquirer will then be replaced carefully by synonyms (through WordNet) and ConceptNet will "filter out" expressions not related to the examined database. For a further "filter out" process, we further propose the use of bigrams of Adjective Noun Pairs instead of simply using adjectives or nouns separately to take advantage of an adjective with a strong sentiment instead of using a neutral noun. Thus, using language models and natural language processing for semantic analysis of multimedia tags the results are expected to improve[44]. For analyzing further the affect in language, the ANET and ANEW dictionaries [2] could be also investigated.

With this architecture, we expect to enrich and thus visualize better a number of well-known Psychological Foundations such as the Circumplex of Affect[45], the Plutchik's Wheel of Emotions[46] and the Geneva Wheel[47] with sentiment values that are mapped with words defining context, having these words spread at one of the four quadrants[48]. Thus,

[2] http://csea.phhp.ufl.edu/Media.html

enriching the gamut of polarized emotions, the representation of emotion intensity will be allowed, as well as the similarity of contrast between various emotions categories and the associated words defining context

Currently, due to the fact that the detected set of (visual) context concepts are expected to operate as mid-level representations and infer automatically the sentiments (the "Why" of social media), we are investigating whether these representations could be used to estimate appraisals and shed light into the first mapping scheme of the componential model which still remains unexplored[19]. Further research on this direction might be able to show whether the combination of this arguments is feasible and can introduce a new data acquisition protocol suitable for context.

Thus, it is important to find comprehensive and thorough answers to the currently unaddressed questions posed above, and fully explore the terrain of context in Affective Computing shedding light in each of its subcomponents. More specifically, exploring the terrain of the modeling of the cognitive theories of context based on affective interaction, the extraction of context information, the incorporation of contextual information in affective corpora focusing on how it should be represented, what contextual information is relevant, as well as how the integration of contextual information, would improve the performance of multimodal frameworks. Such knowledge is expected to enable several exciting directions for further investigation. It could also enable technologies such as context based and affect-aware intelligent tutors, human-embodied conversational agent interactions, independent living and personal wellness technologies, broadcast video news technologies, face recognition in personal photos in the wild, face recognition systems that utilize body and clothing, and educational tools. Moreover, in the area of arts various applications, using the term context, range from analysis of aesthetics, to arts installations. Similarly, in the area of clinical applications context aware technologies such as depressions severity detection, stress/pain motoring, emotion related disorders such as autism etc. could be developed. Such technologies would have large impact in domains such as arts installations and entertainment, education (e.g. gaming applications) and learning styles and healthcare applications.

Acknowledgements

Current research work has been partially funded by the Greek Ministry of Education, Education and Lifelong Learning program, project code MIS 380328 "Cross-disciplinary research in Affective Computing for physiology and activity recognition in assistive environments - STHENOS".

References

1. Douglas-Cowie, E., Cowie, R., Cox, C., Amier, N., Heylen, D.. The sensitive artificial listener: an induction technique for generating emotionally coloured conversation. In: *LREC Workshop on Corpora for Research on Emotion and Affect*. European Language Resources Association(ELRA); 2008, p. 1–4.
2. Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., Paiva, A.. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: *6th ACM International Conference on Human-Robot Interaction (HRI)*. IEEE; 2011, p. 305–311.
3. Picard, R.. Future affective technology for autism and emotion communication. *Philosophical transactions of the Royal Society of London Series B, Biological Sciences* 2009;**364**(1535):3575–3584.
4. Littlewort, G.C., Bartlett, M.S., Salamanca, L.P., Reilly, J.. Automated measurement of children's facial expressions during problem solving tasks. In: *International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*. IEEE; 2011, p. 30–35.
5. Hammal, Z., Suarez, M.T.. Towards context based affective computing. In: *Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE; 2013, p. 802–802.
6. Schilit, B., Adams, N., Want, R.. Context-aware computing applications. In: *Proceedings of the First Workshop on Mobile Computing Systems and Applications*. IEEE; 1994, p. 85–90.
7. Brown, P.J., Bovey, J.D., Chen, X.. Context-aware applications: from the laboratory to the marketplace. *Personal Communications* 1997; **4**(5):58–64.
8. Ryan, N.S., Pascoe, J., Morse, D.R.. Enhanced reality fieldwork: the context-aware archaeological assistant. In: *Computer Applications in Archaeology*; British Archaeological Reports. Tempus Reparatum; 1998, p. 182–196.
9. Brown, P.J.. The stick-e document: a framework for creating context-aware applications. In: *Proceedings of EP'96, Palo Alto*. 1996, p. 182–196.
10. Franklin, D., Flaschbart, J.. All gadget and no representation makes jack a dull environment. In: *Proceedings of the AAAI Spring Symposium on Intelligent Environments*. 1998, p. 155–160.
11. Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M., Steggles, P.. Towards a better understanding of context and context-awareness. In: *Handheld and Ubiquitous Computing*. Springer; 1999, p. 304–307.
12. El Kaliouby, R., Robinson, P., Keates, S.. Temporal context and the recognition of emotion from facial expression. In: *Proceedings of the HCI International Conference*. American Psychological Association; 2003, p. 631–635.

13. Knudsen, H.R., Muzekari, L.H.. The effects of verbal statements of context on facial expressions of emotion. *Journal of Nonverbal Behavior* 1983;**7**(4):202–212.
14. Bonin, F., Bock, R., Campbell, N.. How do we react to context? annotation of individual and group engagement in a video corpus. In: *Privacy, Security, Risk and Trust (PASSAT), International Conference on Social Computing (SocialCom)*. IEEE; 2012, p. 899–903.
15. Bock, R., Wendemuth, A., Gluge, S., Siegert, I.. Annotation and classification of changes of involvement in group conversation. In: *Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE; 2013, p. 803–808.
16. Duric, Z., Gray, W.D., Heishman, R., Li, F., Rosenfeld, A., Schoelles, M.J., et al. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. *Proceedings of the IEEE* 2002;**90**(7):1272–1289.
17. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2009;**31**(1):39–58.
18. Calvo, R.A., D'Mello, S.. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing* 2010;**1**(1):18–37.
19. Mortillaro, M., Meuleman, B., Scherer, K.R.. Advocating a componential appraisal model to guide emotion recognition. *International Journal of Synthetic Emotions (IJSE)* 2012;**3**(1):18–32.
20. McKeown, G., Valstar, M., Cowie, R., Pantic, M., Schroder, M.. The semaine database: annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing* 2012;**3**(1):5–17.
21. Siegert, I., Bck, R., Wendemuth, A.. The influence of context knowledge for multi-modal affective annotation. In: *Human-Computer Interaction. Towards Intelligent and Implicit Interaction*; vol. 8008 of *Lecture Notes in Computer Science*. Springer; 2013, p. 381–390.
22. Lee, C.C., Busso, C., Lee, S., Narayanan, S.S.. Modeling mutual influence of interlocutor emotion states in dyadic spoken interactions. In: *Proceedings of Interspeech*. ISCA; 2009, p. 1983–1986.
23. Graves, A., Fernández, S., Schmidhuber, J.. Bidirectional lstm networks for improved phoneme classification and recognition. In: *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications - Volume Part II*; ICANN'05. Springer-Verlag; 2005, p. 799–804.
24. Kapoor, A., Burleson, W., Picard, R.W.. Automatic prediction of frustration. *International Journal of Human-Computer Studies* 2007; **65**(8):724–736.
25. Pantic, M., Bartlett, M.. Machine analysis of facial expressions. In: *Face Recognition*. I-Tech Education and Publishing; 2007, p. 377–416.
26. Hammal, Z., Kunz, M.. Pain monitoring: A dynamic and context-sensitive system. *Pattern Recognition* 2012;**45**(4):1265–1280.
27. Hammal, Z., Kunz, M., Arguin, M., Gosselin, F.. Spontaneous pain expression recognition in video sequences. In: *Proceedings of the international conference on Visions of Computer Science*. British Computer Society; 2008, p. 191–210.
28. Pantic, M.. Machine analysis of facial behaviour: Naturalistic and dynamic behaviour. *Philosophical Transactions of The Royal Society B: Biological sciences* 2009;**364**(1535):3505–3513.
29. Dhall, A., Goecke, R., Lucey, S., Gedeon, T.. Collecting large, richly annotated facial-expression databases from movies. *IEEE MultiMedia* 2012;**19**(3):0034.
30. Russell, J.A.. Core affect and the psychological construction of emotion. *Psychological review* 2003;**110**(1):145.
31. Curhan, J.R., Pentland, A.. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology* 2007;**92**(3):802.
32. McCowan, I., Gatica-Perez, D., Bengio, S., Lathoud, G., Barnard, M., Zhang, D.. Automatic analysis of multimodal group actions in meetings. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2005;**27**(3):305–317.
33. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.. Imagenet: A large-scale hierarchical image database. In: *Conference on Computer Vision and Pattern Recognition, (CVPR)*. IEEE; 2009, p. 248–255.
34. Jiang, Y.G., Ye, G., Chang, S.F., Ellis, D., Loui, A.C.. Consumer video understanding: A benchmark database and an evaluation of human and machine performance. In: *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*. ACM; 2011, p. 29.
35. Strassel, S., Morris, A., Fiscus, J.G., Caruso, C., Lee, H., Over, P., et al. Creating havic: Heterogeneous audio visual internet collection. In: *LREC*. European Language Resources Association (ELRA); 2012, p. 2573–2577.
36. Kennedy, L.S., Chang, S.F., Kozintsev, I.V.. To search or to label?: predicting the performance of search-based automatic image classifiers. In: *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*. ACM; 2006, p. 249–258.
37. Lenat, D.. The dimensions of context-space. *Cycorp Technical Report* 1998;.
38. Kapoor, A., Picard, R.W.. Multimodal affect recognition in learning environments. In: *Proceedings of the 13th annual ACM International Conference on Multimedia*. ACM; 2005, p. 677–682.
39. Arroyo, I., Cooper, D.G., Burleson, W., Woolf, B.P., Muldner, K., Christopherson, R.. Emotion sensors go to school. In: *Proceedings of the international conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*. IOS Press; 2009, p. 17–24.
40. Pu, P., Chen, L., Hu, R.. Evaluating recommender systems from the users perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction* 2012;**22**(4-5):317–355.
41. Stone, P.J., Dunphy, D.C., Smith, M.S., Ogilvie, D.M.. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press; 1966.
42. Miller, G.A.. Wordnet: a lexical database for english. *Communications of the ACM* 1995;**38**(11):39–41.
43. Liu, H., Singh, P.. Conceptnet-a practical commonsense reasoning toolkit. *BT technology journal* 2004;**22**(4):211–226.
44. Soleymani, M., Larson, M., Pun, T., Hanjalic, A.. Corpus development for affective video indexing. *IEEE Transactions on Multimedia* 2014;**16**(4).
45. Russell, J.. A circumplex model of affect. *Journal of Personality and Social Psychology* 1980;**39**(6):1161–1178.
46. Plutchik, R.. *Emotion: A Psychoevolutionary Synthesis*. Harper & Row New York; 1980.
47. Scherer, K.R.. What are emotions? and how can they be measured? *Social Science Information* 2005;**44**(4):695–729.
48. Borth, D., Ji, R., Chen, T., Breuel, T., Chang, S.F.. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: *Proceedings of the 21st ACM International Conference on Multimedia*. ACM; 2013, p. 223–232.