Context in Affective Multiparty and Multimodal Interaction: Why, Which, How and Where?

Aggeliki Vlachostergiou*
Image Video and Multimedia Systems Laboratory
National Technical University of Athens
Iroon Polytexneiou 9, 15780 Zografou, Greece
aggelikivl@image.ntua.gr

George Caridakis†
Image Video and Multimedia Systems Laboratory
National Technical University of Athens
Iroon Polytexneiou 9, 15780 Zografou, Greece
gcari@image.ntua.gr

Stefanos Kollias
Image Video and Multimedia Systems Laboratory
National Technical University of Athens
Iroon Polytexneiou 9, 15780 Zografou, Greece
stefanos@cs.ntua.gr

ABSTRACT

Recent advances in Affective Computing (AC) include research towards automatic analysis of human emotionally enhanced behavior during multiparty interactions within different contextual settings. Current paper delves on how is context incorporated into multiparty and multimodal interaction within the AC framework. Aspects of context incorporation such as importance and motivation for context incorporation, appropriate emotional models, resources of multiparty interactions useful for context analysis, context as another modality in multimodal AC and context-aware AC systems are addressed as research questions reviewing the current state-of-the-art in the research field. Challenges that arise from the incorporation of context are identified and discussed in order to foresee future research directions in the domain. Finally, we propose a context incorporation architecture into affect-aware systems with multiparty interaction including detection and extraction of semantic context concepts, enhancing emotional models with context information and context concept representation in appraisal estimation.

Categories and Subject Descriptors

A.1 [**General Literature**]: Introductory and Survey; H.1.2 [**User/Machine Systems**]: Human information processing; H.5.1 [**Multimedia Information Systems**]: Evaluation/methodology; H.5.3 [**Group and Organization Interfaces**]: Evaluation/methodology; J.4 [**SOCIAL AND BEHAVIORAL SCIENCES**]: Psychology

*Corresponding author.

†George Caridakis is also affiliated with the Department of Cultural Technology and Communication, University of the Aegean.

UM3I'14, November 16, 2014, Istanbul, Turkey.

ACM 978-1-4503-0652-2/14/11 ...$15.00.
http://dx.doi.org/10.1145/2666242.2666245.

Keywords

Affective Computing, Multiparty and Multimodal Interaction, Context Modeling, Context-Aware Systems, Adaptive Systems

1. WHY IS THE CONTEXT IN AFFECT MULTIPARTY INTERACTION EXPLORED?

A multiparty interaction is a communication configuration that involves three or more interlocutors. In the field of dialogue processing for Human-Machine Interfaces (HMI), researchers have tried to build an environment in which the machine interprets an individual human's behaviors and provides suitable contextual information for a more engaging interaction. One of the key technologies necessary for this is the ability to read in terms of understanding individual human behaviors in a robust and improved way via the contextual aspect. In contrast, the processing of multiparty interactions aims to build an environment in which an artificial agent can participate in an ongoing human conversation (i.e., in a multiparty interaction).

However, despite the significant amount of research on automatic affect analysis, the current state-of-the-art has not yet achieved the long-term objective of robust multiparty affect analysis, particularly context-aware affect analysis and interpretation. Indeed, it is well known that affect production is accordingly displayed in a particular context, such as the ongoing task, multiple participants involved in the conversation, their interaction with context, turn-taking patterns, the identity and natural expressiveness of the individual. The context tells us which expressions are more likely to occur and thus can bias the classifier toward the most likely/relevant classes. Without context, even humans may misunderstand the observed facial expression. By tackling the issues of context-aware affect analysis, i.e. careful study of contextual information and its relevance in domain-specific applications, its representation, its modeling and incorporation including its effect on the performance of existing affect recognition methods, we make a step towards the underlying aim of advancing the interaction to a more affective, socially rich, fluent and engaging one.

In recent years, the automatic analysis of multiparty human behavior has been attracting an increasing amount of attention from researchers because of the width of poten-

tial applications and its intrinsic scientific challenges. Many research fields including pervasive and ubiquitous computing, multimodal natural interaction, ambient assisted living, computer supported collaborative work, surveillance etc. have been benefitted by the awareness due to the fact that a system can provide more advanced services to people only if it can understand much more about users' attitudes, personality [37], social relationships [37], their role [29] along with the degree of their engagement and interest during the interaction [2], as well as the knowledge of the general interactional context/situational setting [8, 38]. Additionally, contextual information such as discourse and social situations [5], general situational understanding [10], cultural background [23] could also assist in evaluating situations appropriately.

According to pioneering work introducing the term context-awareness in Computer Science (CS) [30], the important aspects of context are: who you are with, where you are, when, what resources are nearby, what the user is doing and this information is used to determine why the situation is occurring. In a similar approach, in [8] authors define context as location, identities of the people around the user, the time of day, season, temperature, etc. Other approaches [28] include context as the user's location, environment, identity and time while others have simply provided synonyms for context. For a more extended overview on context-awareness in a wider framework, the reader is referred to [1].

To fulfill the need of incorporating context in multiparty human communicative behavior, the context model which was formalized as the combination of the "Identity", "Time", "Location" and "Activity" contextual types presented in [1] was enriched [41] to include an additional contextual type called "Relations", referring to information about any possible relation that a person may establish with others during an interaction (multiparty). The "Relations" type expresses a dependency among the interlocutors that emerges and also acts as indices to other sources of contextual information. For example, given a person's affiliation, related information such as social associations, connections, information about friends, enemies, neighbors, co-workers, relatives, etc. can be acquired. Recently the term Relations has been used to refer to the relation between the individual and the social context in terms of perceived involvement [6] and to the changes detected in a group's involvement in a multiparty interaction [5]. Thus, it is clear that understanding the multiparty human communicative behavior implies understanding the modifications of the social structure and dynamics of small [17] and large groups (friends, colleagues, families, students, etc.) [11] and the changes in individuals' behaviors and attitudes that occur because of their membership in social and situational settings [38].

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 users and the system, with the definition of [14] approaching better the understanding of human affect signals in AC. More specifically, this work [14] summarizes appropriately the key aspects of context with respect to the human communicative behavior in terms of the context-aware formalization: the human-human multiparty interactions, the non-linguistic conversational signal or emotion that is communicated, how the information is communicated (the person's facial expression, head movement, tone of voice, and hand and body gestures), the context under which the information is passed on: where the user is, what his current task is, and how he/she feels and the (re)action that should be taken to satisfy user's needs and requirements. Thus, so far the efforts on human multiparty behavior understanding are usually context independent [26] while the above presented context-aware methodology answers one or more questions separately each time [4, 34]. Overall, further research is needed in answering the above context-aware aspects simultaneously and propose a context formalization framework.

2. WHICH ARE THE APPROPRIATE AFFECT MODELS FOR CONTEXT?

Stressing that the inclusion of affect representation into a framework for multiparty affect analysis is of primary importance, incorporating models and paradigms developed by psychologists for the classification of affective states [2] is a pressing need and remains a challenging issue. Strengthening the connection with affect representation models would allow for the first steps towards the detection of alternative theoretical affect models, cognitive processes, attitudinal states 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 [40]. However, in view of their suitability to context modeling, emphasis is given on 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 [9, 18].

Recently, a research attempt has suggested 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 [25], 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 proposed. 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 [25]. Thus, this latter appraisal based model is more appropriate to the context-aware formalization as it decomposes the appraisal process into the two above layers and deals in a more effective way with the multiparty model's challenges such as the dynamic and bidirectional individual-group relationship and the modeling of context in order to capture more suitable multiparty social setting patterns w.r.t. the contextual aspect.

3. WHICH ARE THE RESOURCES OF MULTIPARTY INTERACTION? DATA AND ANNOTATION

A major concern in studying the context modeling process and its incorporation into multipaty AC frameworks is

the collection and annotation of suitable multipary data.The annotation process generally reveals that context is indeed very difficult to model in naturalistic interaction which usually occurs in social situations, where it is not easy to obtain clear from noise recordings of spontaneous and natural expressions of an individual. Moreover, due to the fact that there is not an agreed data acquisition protocol that could be applied to provide improved results, the task of identifying and extracting contextual information in existing multiparty affective corpora becomes even harder. Even for a human expert is difficult define or identify what constitutes context related to emotion. Recent developments in hardware (e.g. depth cameras) and computer vision techniques have attempted to tackle this problem by introducing more advanced settings and capture technologies (e.g. Microsoft Kinect) [20]. So far, there has been much less emphasis on multiparty dialogue based on context-dependent data that could assist in shaping our understanding of the problem itself and shed light into main annotation problems in terms of incorporating context in multiparty interactions.

Few exemptions that satisfy some of the above requirements are organized according to the different interaction contexts: task-dependent scenario usually for educational or entertainment purposes, group interactions ranging from scripted conversations (e.g. the M4 corpus), to task-based conversations with and without pre-assigned roles (e.g. the AMI and the Mission Survival 2 (MSC-2) corpora respectively), to real-life interactions that would have happened irrespectively of the recording process (e.g. in the NIST corpus) as well as media databases [16, 17]. In the former category representative examples are the three-party multimodal corpora [2] and the Tutorbot Corpus [20] where the whole personality traits acquisition process is driven by the context of a highly engaged quiz game scenario (participants are at the same distance from each other, the game master poses the questions, the team discusses the questions, one team member gives the answers etc.) and during a game solving collaborative dialogue (the two participants involved in a dialogue are paired into teams, the tutor and the two participants are sitting around a table at approximately equal distance from one another, etc.), with the underlying aim of detecting social conversational cues such as engagement, turn taking patterns, level of interest etc. In the final category of datasets, media, radio news and televised political debates are included. However, the main problem with such data is that the audio signal is usually noisy due to the presence of several people. For example, the political debate Canal9 database [36] even though is suitable for extracting social constructs, on the other hand implies difficulties as people tend to move their heads a lot and near-frontal views of the face are not always visible, making the visual data difficult to process even with the state-of-the-art methods.

Thus, it is clear that the domain is still in its early stages and no major efforts have been done yet for the collection of context-dependent data specifically aimed at the analysis of contextual information. Most of the works in the literature use data originally aimed at different purposes and annotated ad hoc for satisfying the needs of the performed experiment each time. Obtaining the ground truth can be very challenging and requires a strict data annotation protocol regarding the global definition of context for the annotators, but also regarding the starting and ending points of such an episode. However, multiparty interactions 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. Without a significant effort for data collection and annotation it will be impossible to tackle solutions to current research issues in the field. For a more extended list of multiparty multimodal corpora the reader is referred to [17].

In terms of current annotation schemes w.r.t. to the contextual aspect in AC frameworks, we are aware of the two following recent annotation attempts [6, 5]. In the former, the authors aim to explore the relation between the individual and the social context in terms of the perceived involvement. Focus is given on the behavior of the four participants in the group context. Their behavior during the free conversation is analyzed first individually and then as a group, annotating the degrees of the participants' engagement and involvement. Discrete labels of "-" (minus) and "+" (plus) have been used to indicate engagement decrease and increase respectively. In the latter work, the authors applied a similar approach for annotating changes in involvement, using three continuous labels ("+"(increasing), "-"(decreasing) and "0" (no change) for the whole session).

4. HOW IS THE CONTEXTUAL INFORMATION INCORPORATED INTO AC?

Following context integration on multiparty affect production in terms of elicitation and context incorporation in emotion corpora, this section reviews context-awareness in multiparty AC, providing a number of works that could assist in establishing context as an additional modality in multiparty AC.

In [19], the authors proposed research for determining automatically the addresse, i.e. the person at whom the speech is directed by combining utterance, gaze and gesturing features with contextual features such as the interaction history, the meeting history, the user and the spatial context; however, this proposed scheme has not been validated. In the work of [38], authors considered situational context in terms of simultaneously modeling speech of all meeting participants by employing bidirectional Long Short-Term Memory (BLSTM) recurrent neural networks to achieve a considerable accuracy compared to previous methods based on HMMs. Moreover, a Dynamic Bayesian Network (DBN) has been presented in [3] to estimate the joint focus of group participants by combining the visual attention with the contextual cue of speaking activity. This work resulted in an improved VFOA (Visual Focus of Attention) recognition from head orientation automatically estimated from a single camera. Additionally, in terms of analyzing the relation between interest and speech modality based on verbal and non-verbal cues, the authors of [39] attempted to detect the highly involvement (amused, disagreeing, other) of interlocutors with the concept of activation in emotion modeling. Additionally, it was reported that the 2% of the utterances, as well as certain non-verbal prosodic features (pitch, energy) were correlated with involved utterances. At a later point, this work, was extended by adding contextual features such as speaker identity and meeting type along with lexical features such as perplexity and utterance length.

Finally, an automatic context-aware system to detect the personality traits of Extraversion and the Locus of Control in

a meeting environment by means of audio and visual features has been implemented in [27]. In this work, the contextual aspect refers to the information about the relational context. The classification task is applied to thin slices of behavior (1-minute sequences), while the performances of several training and testing instance setups have been tested using SVMs, including a restricted set of audio features obtained through feature selection. The outcomes have been proven to outperform considerably over existing results, providing evidence about the feasibility of the multimodal analysis of personality when incorporating the contextual information.

5. HOW CAN WE ADVANCE THE FIELD?

Domains where human behavior understanding is a crucial need (e.g., Human-Computer Interaction (HCI), AC and Social Signal Processing (SSP)) rely on advanced techniques to automatically interpret complex context-aware multiparty patterns generated when humans interact with others. Thus, this is a difficult problem where many issues are still open, and thus further work is needed along this front.

5.1 Challenges

The main criticism of [17] that reviews work on multiparty data is that it is hard to **capture and take into account all possible context-aware social constructs** in affect elicitation (explicitly or implicitly) [32] of spontaneous human behavioral events during multiparty interaction. This is partly due to the limitations in capture and synthesis technologies and the difficulty in modeling certain contextual information such as personality, culture, background, situation, and so on. Therefore, until systems are able to model all the signals that take place during multiparty interaction, developed ECAs in a multiparty human computer interaction will remain unable to mutually influence and affect each other's behavior and internal states (e.g. attitudes) in a host of different ways.

Another issue is related to the **data collection in the wild in terms of naturalistic conditions and its annotation [13]**: Both psychologists and engineers tend to acquire their data in laboratories and artificial settings [12], 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. It would be desirable to collect several corpora of multiparty interactions (groups of various sizes, if possible culture-dependent), under different scenarios. Unfortunately, researchers so far have not come to an agreement for a common annotation method for intragroup interactions due to a number of issues, such as the pre-defined single feed by multiple cameras etc.

Incorporating context as a dimension: This presents particular challenges, as discussed in [21] 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 [25] 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 face-to-face conversational cues: 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 [9].

Addressing **challenges to study and explore the multiparty interactions** in all its multimodal mediums has been attempted in academia. We are aware of a number of research projects concerned with the interpretation of human behavior at a deeper level in group interactions (e.g. AMI/AMIDA project[1]), any social interactions (Social Signal Processing project[2]), as well as the automated understanding of face-to-face social interaction as a research problem in computing (IM2 project[3]).

Contextual design: At which level in the processing stream does contextual information have a role? In the area of multiparty interactions, 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 etc. and user's parameters such as the personality, the person's affective state, the contextual information etc. Thus, for example there is the challenge of incorporating context when designing computational models of social parameters that would be applied on small and large groups in order to capture the interaction process as well as to identify more meaningful features. In other words, more advanced and sophisticated models should be applied [35] in order for the field to move forward.

Evaluating context-aware affect 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 working on modeling intragroup interactions including less than five people [15] implies a number of issues. The main reason for this is the fact that small group interactions are much more challenging, particularly from the technological point of view (e.g. capture difficulties due to the dependencies between roles).

5.2 Proposed Context-aware AC framework

Attempting to recognize the communicative intention including affective and cognitive states of the user and proceed to a more advanced conversation including more affective, socially rich, fluent and engaging social constructs, we take advantage of the similar approaches applied in the multimedia computing area [7] to propose a foundation platform, able of tackling the issues of context-awareness of multiparty communication analysis.

Thus, our proposed architecture is threefold: **a) to detect a set of (visual) semantic concepts from multiparty interactions** that refer to a physical presence of objects or scenes that define context. These visual concepts can be used to fill the affective gap by advancing the proposed mid-level representation to a more expressive ONE and to automatically infer the sentiments (highly subjective

[1] www.amiproject.org
[2] www.sspnet.eu
[3] www.im2.ch

human responses), reflected in the captured video. These lexical concepts can be further used by the publicly available online knowledge sources (OKS) in natural language processing such as the General Inquirer [33], the WordNet [24] and the ConceptNet [22] that contain information about words, concepts, or phrases, as well as connections among 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 an additional "filter out" process, we suggest the investigation 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, to further improve the expected results [31].

b) enrich and thus visualize better a number of well-known Psychological Foundations [18] such as the Circumplex of Affect, the Plutchik's Wheel of Emotions and the Geneva Wheel with sentiment values that are mapped with words defining context, having these words spread at one of the four quadrants [7]. Thus, 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.

c) due to the fact that the detected set of context concepts are expected to operate as mid-level representations, we intend to investigate 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 [25]. Further research on this direction might be able to show whether the combination of this argument is feasible and can introduce a new data acquisition protocol suitable for context.

6. WHERE COULD CONTEXT-AWARE AC FRAMEWORKS BE APPLIED?

Despite the many challenges, it is important to keep in mind the potential of this emerging domain to bring researchers a better understanding of how humans continuously assess and respond to emotional stimuli during a multiparty conversation under different contexts. Such a computation analysis of group interactions could assist in exploring the terrain of context in AC shedding light in each of its subcomponents (context in multiparty affect elicitation, interpretation and analysis).

Such knowledge is expected to enable technologies such as context based and affect-aware intelligent tutors, human-embodied conversational agent interactions and independent living and personal wellness technologies. Additionally, such technologies would have large impact in meeting and teaching situations, evaluations in office environments and domains such as naturalistic HMI that can continuously process a variety of multiparty multimodal information from the users as they unfold, monitor the users' internal state and respond appropriately.

7. ACKNOWLEDGMENTS

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

8. REFERENCES

[1] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles. Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing*, pages 304–307. Springer, 1999.

[2] S. Al Moubayed, J. Edlund, and J. Gustafson. Analysis of gaze and speech patterns in three-party quiz game interaction. In *INTERSPEECH*, pages 1126–1130, 2013.

[3] S. O. Ba and J.-M. Odobez. Multi-party focus of attention recognition in meetings from head pose and multimodal contextual cues. In *International Conference on Acoustics, Speech and Signal Processing,ICASSP*, pages 2221–2224. IEEE, 2008.

[4] S. O. Ba and J.-M. Odobez. Recognizing visual focus of attention from head pose in natural meetings. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(1):16–33, 2009.

[5] R. Bock, A. Wendemuth, S. Gluge, and I. Siegert. Annotation and classification of changes of involvement in group conversation. In *Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 803–808. IEEE, 2013.

[6] F. Bonin, R. Bock, and N. Campbell. 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)*, pages 899–903. IEEE, 2012.

[7] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM International Conference on Multimedia*, pages 223–232. ACM, 2013.

[8] P. J. Brown, J. D. Bovey, and X. Chen. Context-aware applications: from the laboratory to the marketplace. *Personal Communications*, 4(5):58–64, 1997.

[9] R. A. Calvo and S. D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1):18–37, 2010.

[10] J. M. Carroll and J. A. Russell. Do facial expressions signal specific emotions? judging emotion from the face in context. *Journal of Personality and Social Psychology*, 70(2):205, 1996.

[11] T. Choudhury and A. Pentland. Modeling face-to-face communication using the sociometer. 5:3–8, 2003.

[12] J. R. Curhan and A. Pentland. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, 92(3):802, 2007.

[13] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon. Collecting large, richly annotated facial-expression

databases from movies. *IEEE MultiMedia*, 19(3):0034, 2012.

[14] Z. Duric, W. D. Gray, R. Heishman, F. Li, A. Rosenfeld, M. J. Schoelles, C. Schunn, and H. Wechsler. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. *Proc. of the IEEE*, 90(7):1272–1289, 2002.

[15] S. Favre, H. Salamin, J. Dines, and A. Vinciarelli. Role recognition in multiparty recordings using social affiliation networks and discrete distributions. In *Proc. of the 10th international conference on Multimodal interfaces*, pages 29–36. ACM, 2008.

[16] D. Gatica-Perez. Analyzing group interactions in conversations: a review. In *International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 41–46. IEEE, 2006.

[17] D. Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, 27(12):1775–1787, 2009.

[18] H. Gunes and B. Schuller. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2):120–136, 2013.

[19] N. Jovanovic and R. op den Akker. Towards automatic addressee identification in multi-party dialogues. In *Proc. of the 5th SIGdial Workshop on Discourse and Dialogue*, pages 89–92, Pennsylvania, USA, 2004. Association for Computational Linguistics.

[20] M. Koutsombogera, S. A. Moubayed, B. Bollepalli, A. H. Abdelaziz, M. Johansson, J. D. Á. Lopes, J. Novikova, C. Oertel, K. Stefanov, and G. Varol. The tutorbot corpus; a corpus for studying tutoring behaviour in multiparty face-to-face spoken dialogue. In *LREC*, pages 4196–4201, 2014.

[21] D. Lenat. The dimensions of context-space. *Cycorp Technical Report*, 1998.

[22] H. Liu and P. Singh. Conceptnet-a practical commonsense reasoning toolkit. *BT technology journal*, 22(4):211–226, 2004.

[23] T. Masuda, P. C. Ellsworth, B. Mesquita, J. Leu, S. Tanida, and E. Van de Veerdonk. Placing the face in context: cultural differences in the perception of facial emotion. *Journal of personality and social psychology*, 94(3):365, 2008.

[24] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[25] M. Mortillaro, B. Meuleman, and K. R. Scherer. Advocating a componential appraisal model to guide emotion recognition. *International Journal of Synthetic Emotions (IJSE)*, 3(1):18–32, 2012.

[26] M. Pantic, A. Nijholt, A. Pentland, and T. S. Huanag. Human-centred intelligent human? computer interaction (hci^2): how far are we from attaining it? *International Journal of Autonomous and Adaptive Communications Systems*, 1(2):168–187, 2008.

[27] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro. Multimodal recognition of personality traits in social interactions. In *Proc of the 10th international conference on Multimodal interfaces*, pages 53–60. ACM, 2008.

[28] N. S. Ryan, J. Pascoe, and D. R. Morse. Enhanced reality fieldwork: the context-aware archaeological assistant. In V. Gaffney, M. van Leusen, and S. Exxon, editors, *Computer applications in archaeology*, British Archaeological Reports, pages 182–196. Tempus Reparatum, 1998.

[29] H. Salamin and A. Vinciarelli. Automatic role recognition in multiparty conversations: An approach based on turn organization, prosody, and conditional random fields. *IEEE Transactions on Multimedia*, 14(2):338–345, 2012.

[30] B. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *Proc of the First Workshop on Mobile Computing Systems and Applications*, pages 85–90. IEEE, 1994.

[31] M. Soleymani, M. Larson, T. Pun, and A. Hanjalic. Corpus development for affective video indexing. *IEEE Transactions on Multimedia*, 16(4), 2014.

[32] M. Soleymani and M. Pantic. Human-centered implicit tagging: Overview and perspectives. In *International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3304–3309. IEEE, 2012.

[33] P. J. Stone, D. C. Dunphy, M. S. Smith, and D. M. Ogilvie. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press, Cambridge, MA, 1966.

[34] F. Talantzis, A. Pnevmatikakis, and A. G. Constantinides. Audio-visual active speaker tracking in cluttered indoors environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 38(3):799–807, 2008.

[35] A. Vinciarelli. Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. *IEEE Transactions on Multimedia*, 9(6):1215–1226, 2007.

[36] A. Vinciarelli, A. Dielmann, S. Favre, and H. Salamin. Canal9: A database of political debates for analysis of social interactions. In *Proc. of the 3rd International Conference on Affective Computing and Intelligent Interaction,(ACII)*, pages 1–4, 2009.

[37] A. Vinciarelli, H. Salamin, and M. Pantic. Social signal processing: Understanding social interactions through nonverbal behavior analysis. In *Proc. of International Workshop on CVPR for Human Behavior*, pages 42–49. IEEE, 2009.

[38] M. Wöllmer, F. Eyben, B. W. Schuller, and G. Rigoll. Temporal and situational context modeling for improved dominance recognition in meetings. In *Proc. of Interspeech 2012, Portland, Oregon, USA*, pages 350–353. ISCA, 2012.

[39] B. Wrede and E. Shriberg. The relationship between dialogue acts and hot spots in meetings. In *Proc. IEEE Automatic Speech Recognition and Understanding Workshop, (ASRU'03)*, pages 180–185, 2003.

[40] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.

[41] A. Zimmermann, A. Lorenz, and R. Oppermann. An operational definition of context. In *Modeling and Using Context*, pages 558–571. Springer, 2007.