Emotion Recognition Based Preference Modelling in Argumentative Dialogue Systems

Niklas Rach
Institute of Communications Engineering, Ulm University
Ulm, Germany
niklas.rach@uni-ulm.de

Klaus Weber
Institute of Computer Science Augsburg University
Augsburg, Germany
klaus.weber@informatik.uni-augsburg.de

Annalena Aicher
Institute of Communications Engineering, Ulm University
Ulm, Germany
annalena.aicher@uni-ulm.de

Florian Lingenfelser
Institute of Computer Science Augsburg University
Augsburg, Germany
lingenfelser@hcm-lab.de

Elisabeth André
Institute of Computer Science Augsburg University
Augsburg, Germany
andre@informatik.uni-augsburg.de

Wolfgang Minker
Institute of Communications Engineering, Ulm University
Ulm, Germany
wolfgang.minker@uni-ulm.de

***Abstract*—Within this work, we present an approach to model the opinion of a human towards a specific topic in a fine-grained way by using weighted bipolar argumentation graphs. In addition, we discuss how the therefore required rating of related aspects can be collected by means of emotion recognition techniques and discuss an application scenario based on the state-of-the-art Argumentative Dialogue System EVA in which the proposed techniques can be applied.**

***Index Terms*—Argumentative Dialogue Systems, Emotion Recognition, Computational Argumentation**

I. Introduction

Solutions for everyday problems can nowadays be found on the internet, with most of them having an own homepage, streams or channels on multiple platforms and a broad community discussing issues and approaches to solve them. Thus, the internet has become a vital tool to deal with them and the easy access to information from all over the world has changed the way people tackle respective situations. One problem that comes with this possibility is the variety of often contradicting information that has to be processed by the user and it has become impossible to take into account even a fracture of the information available on a certain topic while dealing with the respective problem at the same time. Thus, technologies that provide an intuitive and structured approach to this information are of particular interest. Recent work in Natural Language Processing (NLP) addressed the automatic structuring and mining of content of the Internet. Examples are, among others, the field of Argument Mining [9], [28] and Sentiment Analysis [10]. However, the resulting structured data is not practical for human users. Argumentative Dialogue Systems serve as an interface between these structures and the user, thus providing an incremental and intuitive access to contradicting information [15] on a certain topic. Moreover, the underlying technology has the capacity to support opinion forming and decision making based on the user's personal preferences, knowledge, and existing constraints. Hence, systems of this kind are of high interest for applications in the field of smart home, smart environment and ubiquitous computing in general as most of the above-discussed issues occur in respective scenarios. Moreover, suchlike scenarios usually include physical activities (like cooking) which strengthen the need for efficient, comfortable and often hands-free interaction with the respective technology.

Within this work, we introduce an approach to model user preferences based on the emotional feedback on an argument. In particular, we discuss an application scenario for a modified version of the multi-modal Argumentative Dialogue System EVA [19] in which the user's reaction to presented arguments is monitored by emotion recognition techniques and translated into a preference model based on *weighted bipolar argumentation graphs* (BAG) [20]. The remainder of this paper is as follows: Section II discusses related work from the field of Argumentative Dialogue Systems. Section III introduces the opinion model and the employed emotion recognition techniques. The original Dialogue System and the modification for the herein considered scenario are covered in Section IV whereas Section V closes the work with a brief conclusion and outlook.

II. Background and Related Work

Within this Section, we briefly recall related work from the two major fields this work is related to which are (argumentative) Dialogue Systems and Emotion Recognition.

A. Dialogue Systems

Argumentation is a complex domain for Dialogue Systems and actual implementations of the same have to overcome different barriers [33]. Thus, systems of this kind are comparatively rare. A corpus-based example that avoids a model of the dialogue was presented by [21] whereas [24] introduced a system based on Bipolar Weighted Argumentation Frameworks. Both systems, however, are limited solely on the

978-1-5386-9151-9/19/$31.00 ©2019 IEEE

exchange of arguments, meaning that additional possibilities like questioning the validity of an argument are excluded. On the other hand, [34] and [2] introduced systems based on argument games that allow different moves but on the other hand are restricted to heuristic based strategies. However, none of the above mentioned Argumentative Dialogue System extends the interaction to additional modalities like mimic and gestures (on the system side) or the emotional response of the user as in the herein discussed scenario.

In [13], multimodal signals are employed to assess the quality of a debate presented by the user. In contrast, we rely on the emotional feedback in order to model the user's preferences. Bosma and André [3] used emotions to disambiguate dialogue acts by using data from bio sensors, such as electrocardiography, electromyography, skin conductivity and respiration as additional emotional information along with weighted finite state machines.

B. Emotion Recognition

In literature, there can basically be found three approaches to label users' emotional states [35]. Some of them are focused on classifying emotions, others describe the origin of an emotion. *Categorical models* model emotions as distinct categories, *dimensional models* characterize emotions with respect to dimensions, e.g. valance and arousal and *appraisal models* describe emotions "as valued reactions to emotion-eliciting stimuli". [35]. One of the most famous appraisal models is the so-called OCC (Ortony, Clore, and Collins [12]) model and a very common dimensional model is Russel's circumplex model of affect [25], which we aim to use in this work because it allows us to cope with the diversity of emotions by mapping emotions to discrete emotion classes, such as *positive/high-arousal*, *positive/low-arousal*, *neutral*, *neutral/high-arousal*, *neutral/low-arousal* [36], instead of defining fix emotions.

Wagner et al. [30], Cafaro et al. [4] and Wanner et al. [31] also showed promising results in the combinations of video and audio signals for valence and arousal recognition.

To analyze the required social signals in this work, we make use of the SSI framework (Social-Signal-Interpretation) introduced by Wagner et al. [29], which allows for real-time detection, processing and interpretation of multi-modal sensor data. SSI has already been successfully used for emotion analysis along with a virtual conversational agent by Wanner et al. [31] as well as in a preference adapting joke-telling scenario by means of analyzing the user's visual smile and vocal laugh by Weber et al. [32] and for personalized Human-Robot-Interactions in a story-telling scenario by Ritschel et al. [23].

III. Emotion Recognition based Preference Modelling

This section covers the theoretical foundation of preference modelling as well as the technical scheme for mapping an emotional response into the before mentioned model.

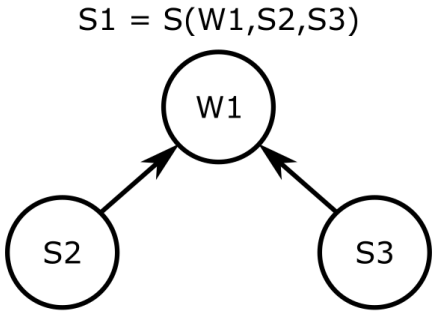

Fig. 1. Sketch of the BAG with s_i the strength of the respective node, w_i the corresponding weight and $s(w_1, s_2, s_3)$ the weight function from Equation 1.

A. Preference Modelling

The preference and opinion of a person on a certain subject may depend on many aspects and different points of view. For example, a person may generally be in favour of pasta dishes but on the other hand aim for a low carbonate diet in the evening. A sophisticated model of the user's opinion thus has to take into account different aspects and still has to be able to picture the personal preference on the overall topic. On the other hand, for topics with a certain complexity, it is not practical to ask the user for a specific rating or judgment on each aspect and especially its effect on previous ones as this is neither efficient nor user-friendly.

The approach proposed within this work is based on argument structures, that encode the dependencies of the different aspects (or arguments) in relations between nodes in a graph. It generally relies on bipolar argumentation structures, meaning that two relations between arguments are possible which are *support* and *attack*. Thus, each argument (node in the graph) is related to another by either backing it up (giving additional reason) or attacking it. We restrict arguments on one relation, which results in a structured tree with the overall claim as root. Thus, it is possible to assign levels to arguments which are basically the connected nodes between it and the root. However, this model only captures the dependencies of arguments and makes no distinction in their strength or validity. In order to model the user's opinion, the herein considered argument structure is extended to a BAG which in addition to the relation assigns a weight w to each node from which its strength can be determined. The strength of an argument i then is a function of its weight w_i and the strength of his child nodes

$$s_i = s(w_i, \{s_j\}) \tag{1}$$

with j a child node of i. Thus, the strength considers both the weight of the argument itself as well as the implications arising from connected arguments. If an argument has no child nodes, its strength equals its weight. For the sake of simplicity and without loss of generality we focus on real-valued strengths and weights between 0 and 1. The basic idea is sketched in Figure 1.

The remaining question is how to adjust the weights for each argument. As mentioned earlier, an argument-wise rating of the user is not practical. Instead, we rely on *preferences* between the arguments, from which we compute a numerical hierarchy that is reflected in the weights. The preferences are determined between arguments that are related to the same parent node, thus allowing to incrementally forming the model.

B. Emotion Recognition

1) General: An intuitive way to judge whether or not a user agrees with an opinion or a single argument is looking at different multi-modal behavioral patterns, such as head nod or gaze. Users, for example, tend to nod if they agree. However, human behavior is far too complex to get an opinion or preference by just looking at one or two of such patterns. More precisely, humans can be excited, bored, happy or disappointed, etc. These affects are part of the well-known circumplex model of affect [25]. And all these affects can describe in a certain way how users think about an opinion. However, taking all these different affects into account for user preference modeling can be challenging and it may be difficult to decide whether a person that has an excited emotional status agrees more than a happy person. This is because people behave very differently and do not always show the same affects - not even to the same extent - for the exact same thing. Introverted people, for example, may not show their emotions as extroverted people do, even if they have the same preference. Therefore, we need a simplification that easily allows us to define whether a user is in favour of an argument or rejects it.

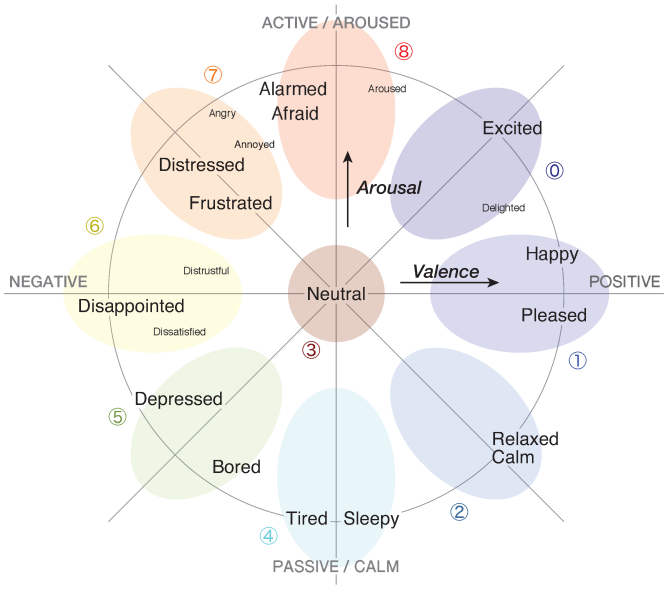

Fig. 2. Two dimensional emotion status model taken from [25].

Matsuda et al. [11] recently proposed mapping the emotional affects of the emotion status model (fig. 2) into a three-class classification task: *Positive*, *Neutral*, *Negative*, i.e. they combine *high-arousal* and *low-arousal* classes.

Using this approach for learning the strength of an argument and therefore the user's preference, we associate *preference classes* with each group:

- *Positive* → *Prefer:* The weight of the argument is increased.
- *Neutral:* The weight of the argument is not modified.
- *Negative* → *Reject:* The weight of the argument is decreased.

The exact value by which the weights are decreased/increased is given by a topic-specific update formula (see Section IV-B).

The presented approach stems on audiovisual input that allows recognizing the two emotional dimensions of the employed valence-arousal model separately, which can afterwards be used to determine the actual aforementioned emotional status.

2) Model Training and Classification: In order to categorize the emotional responses received from the user as reaction to presented arguments, we need classification models to differentiate between *positive*, *neutral* and *negative* reactions. Given the passive (non-speaking) role of the user within our scenario, we rely on the analysis of facial expressions, gestures and postures as source for affective cues that represent the current emotional state.

As a first step towards reliable affect recognition, a descriptive set of modality-specific features needs to be chosen to assess the emotional content within affective channels. We chose the Openface toolbox [37] for the calculation of facial features such as facial orientation, facial landmarks and action units. Especially action units are a commonly used set of features that encode activation of facial musculature and are well suited to distinguish facial expressions and therefore the emotional state of a user. To enrich our observations of emotional responses, we chose to also include body language in our emotion recognition system. Studies like [1] have shown that dynamic body movement as well as static postures actually convey measurable affective information. To perform human pose estimation we apply the open-source software Openpose[1]. With access to an image based estimation of body keypoints and joint configurations we are able to define and recognize the occurrence of postures which convey e.g. a dismissive or open and inviting stance. Furthermore, we calculate features describing the parameters of dynamic gestures, e.g. spatial extent, fluidity or power [8]. As a first attempt to combine the resulting modality feature sets, we aim to merge those features via feature fusion into one global feature set (i.e. super-vector) and train a single, multi-modal classifier. Studies like [11] show that this simple fusion approach is able to improve affect recognition performance compared to uni-modal classification systems. Exploration of more sophisticated fusion strategies at the decision level [16] are planned as future improvements.

Based on these features, we are able to train classification models - given a sufficient amount of annotated training data.

[1]https://github.com/CMU-Perceptual-Computing-Lab/openpose

We plan to incorporate several publicly available emotional corpora such as the Semaine [17] or the Recola [22] database. The recordings in question consist of audiovisual data and feature time-continuous valence annotations. Since our analysis is based on video frames and we defined a valence related three-class classification problem (*positive*, *neutral* and *negative*), we need to discretize and map the valence scores at any given frame-step onto one of our discrete class labels.

Our final affect recognition system aims at online analysis of user responses. Therefore, feature extraction components as well as trained classification models need to be capable of real-time processing and decision making. Support Vector Machines [5] offer a well tested and resource efficient approach to automated class discrimination based on descriptive features and we will apply this classification model before starting experiments with neural network based approaches. Offline classifier training as well as real-time analysis pipelines for online sensor input, feature extraction and classification are realized with the open-source Social Signal Interpretation (SSI) framework [29].

IV. EVA

Recently, the multi-modal Argumentative Dialogue System EVA was introduced which allows a user to discuss controversial topics with a virtual avatar that presents his responses via speech, mimic and gestures [19]. Within this section, we briefly recall the basic modules of the system and discuss the modifications that were made in the context of the present scenario.

A. Original System

In the original system, the user can select his answers from a provided list of allowed utterances which are determined by the underlying dialogue model. This model is based on the dialogue game for argumentation introduced by Prakken [14] which ensures a logically consistent interaction. The respective agent strategy is either based on probabilistic rules or learned by means of multi-agent Reinforcement Learning [18]. The arguments that are employed during the dialogue are structured based on the argument mining scheme of Stab [27] and encoded in an OWL file. Generally, argumentative data that can be mapped into the respective scheme can also be processed by the system and discussed in an own dialogue. The Charamel(TM) avatar which employs the Nuance TTS and all Amazon Polly Voices serves as virtual agent of the system and presents the system utterances to the user. The respective paraphrasing is done by a template based Natural Language Generation (NLG) that utilizes the annotated argument sentence in the OWL file as foundation. A picture of the interface including avatar, response and dialogue history is shown in Figure 3.

B. Virtual Discussion

Within this work, the focus lies on the emotional response of the user to arguments that are presented to him or her. In order to focus on this aspect only, the user's role in the argumentative dialogue is covered by a second agent. Thereby, a virtual discussion between two agents is generated, that the user attends as audience. In doing so, the user is able to review different (pro and con) aspects on a certain topic which would otherwise require the reading of multiple articles and the exploration of additional sources. Hence, the system not only allows to monitor the user's reaction to certain arguments but also provides an intuitive and time efficient interface to arguments on controversial topics.

Fig. 3. Screen capture of the EVA interface including a drop-down menu with possible answers, avatar and dialogue history.

In the scope of such a virtual interaction, each agent takes a stance and tries to win the discussion in terms of the argument game. The user's emotional response can then be utilized to gain insight into the respective preferences. We distinguish three classes of user preference that correspond to the emotion classes discussed in Section III-B: *In favour* (*positive*), *neutral* and *reject* (*negative*). At the beginning of the discussion, each argument in the user model has the same weight 0.5 that corresponds to the preference *neutral*. Whenever an argument is presented by one of the avatars, the emotional signal is used to determine the respective preference class as discussed in Section III-B and the weight of the argument is adjusted accordingly. The update formula for this scenario and an increase of the weight is defined as

$$w_i^{t+1} = w_i^t + \alpha_p(1 - w_i^t) \tag{2}$$

whereas the formula for a decrease is defined as

$$w_i^{t+1} = \alpha_r w_i^t \tag{3}$$

with t the temporal identifier of the corresponding move in the dialogue game and $\alpha_p, \alpha_r \in [0, 1]$ weighting factors.

At the end of each dialogue, the strength function of Equation 1 is utilized to process the new weights through the argument tree and thus to generate the final user model.

V. Conclusion

In this work, we discussed how state of the art emotion recognition techniques can be applied in order to implicitly obtain preferences on certain aspects of a topic from a human user. To this end, we introduced a scheme to map emotional response on the valence-arousal scale into an opinion model based on BAGs. As an example, we introduced a modified version of the Argumentative Dialogue System EVA in which two virtual agents discuss a specific topic in order to convince the audience (the user) of their stance. Future work will mainly aim at different evaluation scenarios for the introduced model in order to determine the respective informative value. One approach will be based on a user survey in which the modelled opinion is compared to the subjective opinion of the user. In addition, a comparison of the emotion recognition based model with explicitly stated preferences will be performed. Finally, an inclusion of the herein presented approach into a working recommendation system is desired.

VI. Acknowledgment

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) within the project "How to Win Arguments - Empowering Virtual Agents to Improve their Persuasiveness", Grant Number 376696351, as part of the Priority Program "Robust Argumentation Machines (RATIO)" (SPP-1999).

References

[1] Atkinson A., Tunstall M., and Dittrich W. Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. Cognition, vol. 104, no. 1, 2007, 5972.

[2] Bench-Capon, Trevor JM. "Specification and implementation of Toulmin dialogue game." Proceedings of JURIX. Vol. 98. 1998

[3] Bosma, Wauter and André, Elisabeth. "Exploiting Emotions to Disambiguate Dialogue Acts." 2014, Proceedings of the 9th international conference on Intelligent user interfaces, 85-92.

[4] Cafaro, Angelo and Wagner, Johannes and Baur, Tobias and Dermouche, Soumia and Torres Torres, Mercedes and Pelachaud, Catherine and André, Elisabeth and Valstar, Michel. 2017. "The NoXi Database: Multimodal Recordings of Mediated Novice-expert Interactions." Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017.

[5] Chang C. and Lin C. LIBSVM: A library for support vector machines. Intelligent Systems and Technology, vol. 2, 2011.

[6] Ekman, P., and Friesen, W. V. (1971). "Constants across cultures in the face and emotion." Journal of Personality and Social Psychology, 17(2), 124-129.

[7] Harreé, Rom. "The social construction of emotions." Blackwell, 1986.

[8] Hartmann B., Mancini B., Buisine S., and Pelachaud C. Design and evaluation of expressive gesture synthesis for embodied conversational agents. in 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2005), July 25-29, 2005, Utrecht, The Netherlands, 2005, 10951096.

[9] Lippi, Marco, and Paolo Torroni. "Argumentation mining: State of the art and emerging trends." ACM Transactions on Internet Technology (TOIT) 16.2 (2016): 10.

[10] Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5.1 (2012): 1-167.

[11] Matsuda, Yuki et al. "EmoTour: Estimating Emotion and Satisfaction of Users Based on Behavioral Cues and Audiovisual Data", Sensors 18.11 (2018): 3978.

[12] Ortony, A., Clore, G., and Collins, A. (1988). "The Cognitive Structure of Emotions." Cambridge university press, 1990.

[13] Petukhova, Volha, et al. "Virtual debate coach design: assessing multimodal argumentation performance." Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017.

[14] Prakken, Henry. "On dialogue systems with speech acts, arguments, and counterarguments." European Workshop on Logics in Artificial Intelligence. Springer, Berlin, Heidelberg, 2000.

[15] Rach, Niklas, et al. "Utilizing argument mining techniques for argumentative dialogue systems." Proceedings of the 9th International Workshop On Spoken Dialogue Systems (IWSDS). 2018.

[16] Lingenfelser F., Wagner J., and Andr E., A systematic discussion of fusion techniques for multi-modal affect recognition tasks. in Proceedings of the 13th International Conference on Multimodal Interfaces, ICMI 2011, Alicante, Spain, 2011, 1926.

[17] McKeown, G., Valstar, M.F., Cowie, R., Pantic, M., Schroeder, M. "The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent." IEEE transactions on affective computing, Vol. 3, No. 1, 2012, 5-17.

[18] Rach, Niklas, Minker, Wolfgang, and Ultes, Stefan. "Markov Games for Persuasive Dialogue." Computational Models of Argument: Proceedings of COMMA 2018 305 (2018): 213.

[19] Rach, Niklas and Weber, Klaus and Pragst, Louisa and André, Elisabeth and Minker, Wolfgang and Ultes, Stefan "EVA: A Multimodal Argumentative Dialogue System." Proceedings of the 2018 on International Conference on Multimodal Interaction. ACM, 2018.

[20] Amgoud, Leila, and Jonathan Ben-Naim. "Evaluation of arguments in weighted bipolar graphs." International Journal of Approximate Reasoning (2018).

[21] Rakshit, Geetanjali, et al. "Debbie, the debate bot of the future." Advanced Social Interaction with Agents. Springer, Cham, 2019. 45-52.

[22] Ringeval F., Sonderegger A., Sauer J., and Lalanne D. Introducing the recola multimodal corpus of remote collaborative and affective interactions, in IEEE FG, 2013, 18.

[23] Ritschel, Hannes and Baur, Tobias and André, Elisabeth. "Adapting a Robot's linguistic style based on socially-aware reinforcement learning." In Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 2018.

[24] Rosenfeld, Ariel, and Sarit Kraus. "Strategical Argumentative Agent for Human Persuasion." ECAI. Vol. 16. 2016.

[25] Russell, J. A. "A circumplex model of affect." Journal of Personality and Social Psychology, 39(6), 1161-1178. 1980.

[26] Scherer, Klaus R., and Marcel R. Zentner. "Emotional effects of music: Production rules." Music and emotion: Theory and research 361 (2001): 392. 2001

[27] Stab, Christian, and Iryna Gurevych. "Parsing argumentation structures in persuasive essays." Computational Linguistics 43.3 (2017): 619-659.

[28] Stab, Christian, Tristan Miller, and Iryna Gurevych. "Cross-topic Argument Mining from Heterogeneous Sources Using Attention-based Neural Networks." arXiv preprint arXiv:1802.05758 (2018).

[29] Wagner, Johannes and Baur, Tobias and Damian, Ionut and Kistler, Felix and André, Elisabeth, "The Social Signal Interpretation Framework (SSI) Framework." Proceedings of the 21st ACM International Conference on Multimedia, 2013, 831-834.

[30] Wagner, Johannes and Lingenfelser, Florian and Bee, Nikolaus and André, Elisabeth. "A Social Signal Interpretation (SSI) - A Framework for Real-time Sensing of Affective and Social Signals." KI - Knstliche Intelligenz, Springer Berlin/Heidelberg (2011): 251-256.

[31] Wanner, Leo, et al. "Kristina: A knowledge-based virtual conversation agent." International Conference on Practical Applications of Agents and Multi-Agent Systems. Springer, Cham, 2017.

[32] Weber, Klaus and Ritschel, Hannes and Aslan, Ilhan and Lingenfelser, Florian and André, Elisabeth. 2018. "How to Shape the Humor of a Robot - Social Behavior Adaptation Based on Reinforcement Learning." In Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18). ACM, New York, NY, USA, 154-162. DOI: https://doi.org/10.1145/3242969.3242976

[33] Yuan, Tangming, et al. "Informal logic dialogue games in humancomputer dialogue." The Knowledge Engineering Review 26.2 (2011): 159-174.

[34] Yuan, Tangming, David Moore, and Alec Grierson. "A human-computer dialogue system for educational debate: A computational dialectics approach." International Journal of Artificial Intelligence in Education 18.1 (2008): 3-26.

[35] André, Elisabeth. "Experimental methodology in emotion-oriented computing." IEEE Pervasive Computing 10.3 (2011): 54-57.

[36] Vogt, Thurid and André, Elisabeth and Wagner, Johannes and Gilroy, Steve and Charles, Fred and Cavazza, Marc. "EReal-time vocal emotion recognition in artistic installations and interactive storytelling: Experiences and lessons learnt from CALLAS and IRIS." 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (2009): 1-8.

[37] B. Amos, B. Ludwiczuk, M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications." CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.