Paradigms for the Construction and Annotation of Emotional Corpora for Real-world Human-Computer-Interaction

Markus Kächele[1], Stefanie Rukavina[2], Günther Palm[1], Friedhelm Schwenker[1] and Martin Schels[1]
[1]*Institute of Neural Information Processing, Ulm University, 89069 Ulm, Germany*
[2]*Department of Psychosomatic Medicine and Psychotherapy, Medical Psychology, Ulm University, 89075 Ulm, Germany*

Keywords: Affective Computing, Affective Corpora, Annotation.

Abstract: A major building block for the construction of reliable statistical classifiers in the context of affective human-computer interaction is the collection of training samples that appropriately reflect the complex nature of the desired patterns. This is especially in this application a non-trivial issue as, even though it is easily agreeable that emotional patterns should be incorporated in future computer operating, it is by far not clear how it should be realized. There are still open questions such as which types of emotional patterns to consider together with their degree of helpfulness for computer interactions and the more fundamental question on what emotions do actually occur in this context. In this paper we start by reviewing existing corpora and the respective techniques for the generation of emotional contents and further try to motivate and establish approaches that enable to gather, identify and categorize patterns of human-computer interaction.

1 INTRODUCTION

Until now, human-computer interactions are still mainly bound to strict inquiry-response protocols that are generally worked off using mouse and keyboard devices or dialog strategies that are emulating these procedures. However the increase of computational power and the increasing penetration of technical devices in the everyday life makes it appealing to implement cognitive capabilities that are capable of enriching the human-computer dialog towards more intuitive interactions. The main means to approach this goal scientifically from both, the psychological and also the technical perspective (pattern recognition), is to create affective corpora that allow to design and evaluate statistical classifiers for user states. These user states and their explicit definitions should be defined prior to the experimental design of every affective corpus, because they represent the ground truth and the labels needed for classification. However, there is still a massive discussion about the definitions of emotions, dispositions and affective states in general (Hamann, 2012; Lindquist et al., 2013; Scherer, 2005). To simplify this definition clutter, we refer to the term of emotional stimuli including emotional and dispositional states in Section 3. A lot of effort has already been put into this and a variety of corpora that were created under a manifold of different paradigms exist (Walter et al., 2011; Rösner et al., 2012; Valstar et al., 2013; Wöllmer et al., 2013). They differ not only in their work definition of emotions but also in their focus on multi-modality, including speech within the interaction, video analyses and physiological recordings.

In this position paper, we briefly discuss the previously created data collections of affective human-computer interaction and the respective labeling techniques. Based on these experiences we establish requirements for a new recording and annotation framework from a pattern-recognition perspective that accounts for pitfalls and artifacts that arise typically in the collection of data collections that should reflect complex and subtle patterns in a real world scenario.

2 STATUS QUO

2.1 Related Work

The research history of affective computing is heavily influenced by the fields of machine learning and psychology. In the middle of the last century, researchers tried to form a comprehensive theory of human emotions by suggesting a number of discrete *basic emotions* that were thought to cover the whole

Kächele M., Rukavina S., Palm G., Schwenker F. and Schels M..
Paradigms for the Construction and Annotation of Emotional Corpora for Real-world Human-Computer-Interaction.
DOI: 10.5220/0005282703670373
In *Proceedings of the International Conference on Pattern Recognition Applications and Methods* (ICPRAM-2015), pages 367-373
ISBN: 978-989-758-076-5

emotional spectrum (Ekman et al., 1969). Inspired by those categories, in the beginning many researchers proposed prototypical corpora e.g. for facial expressions (Kanade et al., 2000), speech (Burkhardt et al., 2005) or physiological measurements (Healey, 2000) in which small sequences of affective material were recorded and labeled with categorical values such as happiness, sadness or surprise. Soon however it was shown that fixed emotional categories were not suitable for every scenario, with the true range of emotional variety being much more complex than that. Combinations of different emotions and different strengths could not be ignored. Inspired by the drawback of fixed emotional categories and coexisting for a long time, (Russell and Mehrabian, 1977) suggested to partition the emotional space into the three dimensions valence, arousal and dominance (VAD). Consequently Ekman's emotions have been reduced to points and octants in this space.

The first challenge that had to be tackled in the design of a new corpus was the question of how to create ground truth information. All the mentioned corpora have in common that the emotional expressions were acted and recorded in strictly controlled areas. A corpus that contains spontaneous (i.e. non-acted) expressions needs an induction process that is able to guide the participants into the desired octants of the VAD space and thus make the participants feel the emotions rather than faking them. One way to solve this challenge is to accurately design experimental sections that use specific predefined stimuli to induce the desired states. This procedure is non-trivial as it needs substantial psychological knowledge in order to design suitable stimuli. Corpora that rely on this method are the Emo-Rec2 (Walter et al., 2011), in which 6 different VAD octants are induced and the Last Minute corpus (Rösner et al., 2012) that relies on a surprise effect to change from the low-arousal half-space of the VAD space to the high-arousal one. Another possibility for the design stage of a corpus is to decouple recording and the creation of ground truth information by subsequent annotation. By doing this the design process is somewhat liberated from an over-complex induction process and instead focus can be set on providing natural interaction sequences between participants and either other people or an HCI interface. The annotation process demands more effort in this case because the recordings have to be manually annotated for affective reactions. Corpora that are based on this paradigm are the AVEC 2011/2012 corpora (Schuller et al., 2011) (HCI), the AVEC 2013/2014 (Valstar et al., 2013) (HCI), the PIT corpus (Strauss et al., 2008) (both HCI and HHI) and the MHI-Mimicry (Sun et al., 2011) (HHI). The results of machine learning algorithms on these datasets are generally worse than on the acted datasets because the reactions are much more rare and subtle in comparison to the overacted ones. To make recognition more realistic, the non-acted corpora generally consist of more than a single modality as in real-life humans also use multi-modal cues to recognize emotions. Some recent works on unconstrained emotion recognition from multi-modal data comprises (Wöllmer et al., 2013; Glodek et al., 2013; Schels et al., 2014; Kächele et al., 2014; Schels et al., 2013b; Schels et al., 2013a).

2.2 Challenges from a Pattern Recognition View

The successful recognition of affective material in human computer interaction scenarios heavily relies on dealing with the following problems:

- It is very difficult (maybe even impossible) to gather the correct ground truth of emotional sequences because the state of a human can not be accurately measured from the outside, or even worse: many people do have problems to describe and rate their emotional feelings. This is however considered as a psychological phenomenon ("alexithymia") and should be controlled.
- Emotional events occur only rarely. This implies that corpora usually contain only few emotional responses and from a pattern recognition perspective, the classification problem may be highly imbalanced (Thiam et al., 2014).
- Strength and manifestation vary highly across subjects. Binary classification tasks therefore shift to fuzzy multi-class problems (Schwenker et al., 2014).

3 DESIGN OF EMOTIONAL CORPORA

The conception of corpora that contain affective events of sufficient quality and quantity can be very demanding and a large number of constraints have to be satisfied in order to achieve this goal.

3.1 The Human Factor

The single most crucial point in the experimental design and subsequent recording stages is the human factor. A poorly designed experiment in which the subject feels like a foreign body will rarely lead to interesting findings other than the obvious fact that the experimental process bothered the subject. The experimental design should thus have a strong focus on the human factor in the beginning of the concept phase, including the questions *"how should the subjects be like?"* and *"what is necessary to sufficiently motivate/fascinate them?"*. Those questions should be deliberated before designing the actual experiment. Further considerations should include subject specifications like age, gender and experience with technical systems. These considerations are not only important to check for in general for the experimental design, but also because gender for example is shown to have an impact on classification processes and results (Rukavina et al., 2013). This issue should be specifically addressed in further experiments.

3.2 Emotional Stimulation

The induction of affective states can be achieved by various means of which, depending on the context at hand, some may be more suitable than others. In the literature the spectrum of elicitation procedures ranges from external passive consumption of rated affective material like pictures (Lang et al., 2005), films (Codispoti et al., 2008; Hewig et al., 2005; Kreibig et al., 2007) and music (Daly et al., 2014; Nater et al., 2006) in contrast to internal autobiographical induction methods (Labouvie-Vief et al., 2003) to more active ways like the interaction with a computer system (classical HCI) via different modalities such as touch or speech or simply the interaction with other humans. It is common to set subjects a task that should be solved either cooperatively with the help of an HCI system or by operating it (or combinations thereof). Stimuli are presented in the form of feedback be it explicitly by praising or dispraising the subject or implicitly by time-delays or malfunctioning. While posing a task is a reasonable way to ensure that interaction is taking place, the characteristic of this task and the associated stimuli are crucial for the success of the induction process.

Immersion. The grade of immersion of the subjects in the process should be scrutinized before the recording phase begins. Stimuli that look promising on paper such as a fictive story (for example about winning a voyage to an unknown place) and a consequential task that relies on the believability of this story (e.g. now packing a suitcase) can be met with indifference because the affective trigger is the introduction of an unforeseen turn of events (in this example this could be the unknown climate of the destination, which turns out to be arctic instead of subtropical and the subjects packed only bathing suits). The catch is that the subjects know that everything is only fictional and that in reality nothing is at stake (i.e. they will not travel anywhere).

The selection of stimuli should thus be influenced by the expected motivation of the subjects. If the task is to play a game and the reward is a specific amount of real money depending on their performance instead of only a high-score or nothing at all, it can be expected that the subjects will be much more motivated and consequently be more immersed in the setting. Taken together, the intrinsic motivation of the subjects has to be activated.

Personalization. Since affective reactions are highly person dependent, the design should allow easy individual adaptation without destroying the objective. Additionally it might be possible to include user preferences and build up on earlier iterations of the experiment to investigate long time effects. Note, that immersion should not exceed reasonable levels and may never breach ethical boundaries. These individual differences are the results of rating processes of situations/stimuli, which in turn are complex combinations of individual experiences and memories. Therefore, many emotion researchers try to use only standardized emotional material (see above: passive emotion induction) instead of situational phenomena during HCI, trying to minimize the individual rating differences in larger samples. As this will be a critical factor in planning an affective corpus, one should think about self ratings during the experiment (see below). Additionally, by selecting the subjects (and their characteristics) carefully it would be possible to minimize the need for personalization e.g. instead of comparing the reactions of an old lady and a young man during HCI.

Repeatability. The experimental design should be conducted in such a way that the desired emotional events occur per design and not coincidental during interaction. More importantly it should be designed such that it can be repeated several times and but still feels fresh and interesting and not annoying or leading to habituation. A good ratio has to be found between the trials (total amount of emotional inductions) and the effect of falsification over time (with increasing amount of emotional inductions the subjects could show habituation effects, leading to no reaction at all, or the originally desired emotion shifts into another e.g. from boredom to anger). From a pattern recognition perspective it is highly desirable to repeat experiments with the same subjects to make a validation possible and to increase the amount of collected data. A certain degree of randomization of the stimuli can also be beneficial to prevent hidden correlations/dependencies between sequences.

Rating. A manipulation check should be included within the experimental protocol, in terms of an emotional self rating during the interaction. This leads to an explicit label which can be used for classification. In addition, this method helps to exclude individual rating differences. Optimally, this rating process is included at times, where the user needs time to ”recover” from previous tasks or if fulfilling a modular experimental structure, it could be included after finishing one module.

Thus, the choice of specific stimuli used in HCI is crucial, because all these difficulties have to be kept in mind. Stimuli, that rely on a surprise effect (like the one in the voyage example) can only be used once on each subject. The surprise effect will be gone in a second run of the scenario, thus this emotional stimuli would not fulfill all the required characteristics and should be avoided.

All the above mentioned points should be considered accurately during the experimental design stage.

3.3 Annotation Techniques

Since the annotation of affective states is particularly expensive, time consuming and ambiguous, it is mandatory to reconsider the actual annotation techniques and the respective labeling tools. We propose to use a hybrid annotation approach that reflects the inherent progress of the human machine interaction and allows the gradual assignments of categories. The main motivation is that, as discussed before, the affective forms of the user state occur only rarely in the course of a computer interaction (if not planned carefully) and are further only in particular states of interest. Hence the interaction model of a human computer dialog should highlight interesting points of the interaction (either by self rating or during the defined induction sequences), which are then passed to a human expert for labeling. This reduces the workload for annotators dramatically compared to annotating whole interaction sequences and also provides the context for the raters that can be useful for them. This could be the success or failure of a respective sub-task.

The manual annotation is not only useful to confirm or reassign a label that could be used as a prior value for an identified sub-sequence but it should also be used for the inquiry of the strength of a label category or the fuzzy nature of it. This can be achieved by providing a slider that is used to adjust the intensity of a category. Alternatively the variance of multiple raters can also be used to determine a certainty or intensity value. Multiple raters are anyways necessary as the affective state of the user is subjective and thus a “true” state can be approached. One possibility to achieve a high standard of labeled material/sequences could be to have different labelers focusing on different emotional channels (e.g. mimic reactions, speech like sighs and gestures like head nodding) similar to a multiple expert system.

A further issue that must be addressed is the usage of a label tool that does not inflict artifacts or distinct patterns on the annotation that originate from the technical constraints of the procedure (Kächele et al., 2014). An example for such technical pitfalls is using a fixed starting point for a continuous annotation of whole interaction sequences.

3.4 Applications & Categories

(Palm and Glodek, 2013) discussed the topic of human-computer interaction from an affective standpoint by introducing the arguably relevant basic blocks: a set of different applications, where having information about affective user states is actually helpful for the interaction and the respective user states that are informative for the system in a way that measures can be undertaken to improve executions of tasks. Also in (Walter et al., 2013) the detection of negative emotions is determined to be helpful in the context of human-computer interaction.

In (Palm and Glodek, 2013) the authors argue that the computer system should mediate between the user and the actual application when it is unknown to the user and complicated to operate. Important roles or applications in this context occur when a computer behaves as one of the following subjects: *Trainer/teacher*, *monitor*, *organizer*, *consultant* and *servant*.

Further (Palm and Glodek, 2013) argue that it is a system's main objective to maintain a positive attitude towards it and the respective task. Hence it is necessary to recognize negative affective user states to discriminate these against neutral or positive states that mainly signify that everything is going well. The relevant negative categories are thus defined as: *Bored*, *disengaged*, *frustrated*, *helpless*, *over-strained*, *angry* and *impatient*.

4 TOWARDS MORE NATURAL DATA COLLECTIONS

In order to collect realistic data comprising affective human-computer interaction it is arguably necessary to move away from over-simplistic trigger-reaction protocols that are used to elicit emotions. However one has to keep in mind that emotions will be induced after certain events that have occurred during

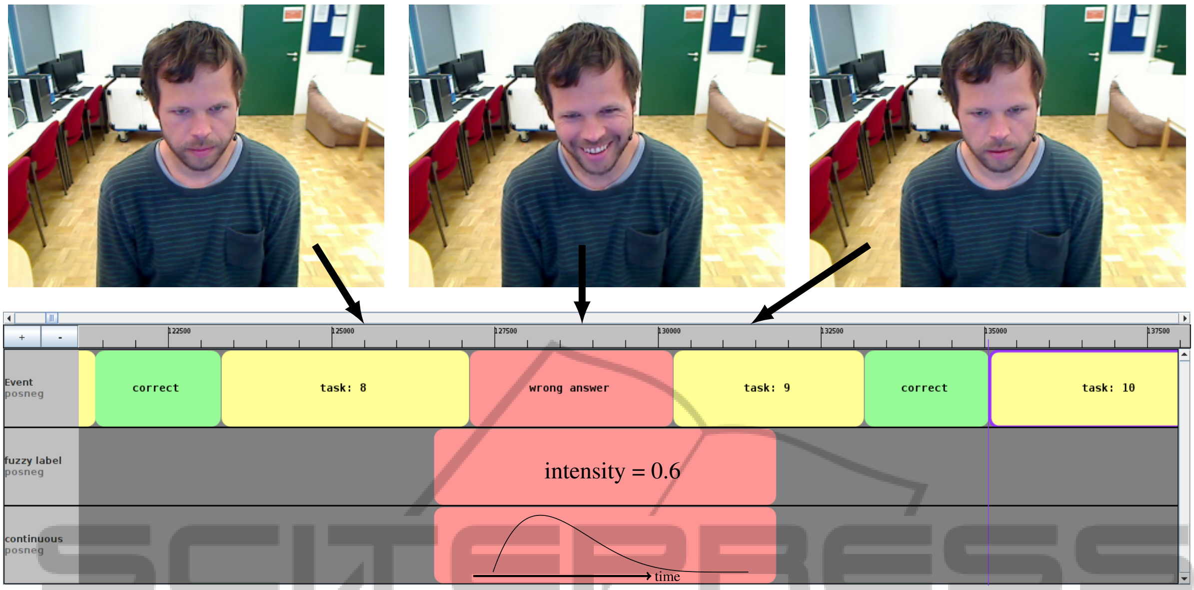

Figure 1: Illustration of the event based hybrid labeling strategy: In a sequence of correctly executed tasks (green) the subject commits an error (red) which marks an interesting spot. This is hence passed to a rater who can annotate the material in the context continuously or using probabilistic values.

the interaction and may have been rated by the subject as positive or negative. Therefore the interesting sequences will have different durations according to the emotion/affective state itself and also according to the intensity. This complicates an automatic analysis and the fusion of different emotional channels, particularly because of their varying latencies which make it even harder. In this section we develop a framework for the construction of an affective corpus for human-computer interaction that reflects the previous considerations and enables one to gather affective patterns that are relevant for this purpose.

For these different applications it is desirable to identify subjects that have an intrinsic motivation for a given task. Intuitive examples for that are students that use a vocabulary trainer in order to improve their language skills and elderly people and patients in a hospital that are monitored for example during the execution of exercises in a rehabilitation action. Thus, a subject conducts a given task interacting with the computer in the teaching domain or monitored by a computer system while all available data is captured in order to create a corpus.

As described earlier these interesting sequences are then passed to human raters that inspect them with respect to whether an affective pattern is present. In order to assign useful labels for this data, the hybrid labeling technique described above should be used in order to reduce the workload of the raters. The interesting points in the different applications can be identified by examining the dialog model of the interaction at the design time. For example success or failure of a given subtask can be very intuitively used as cue for this purpose. In the context of a teaching situation using a vocabulary trainer a wrong answer for a particular word or a wrong pronunciation can be used as a signal for segmentation which is available at no additional cost for this application. Another example is the correct execution of rehabilitation exercises, which can easily be determined by human feedback (e.g. given by the physiotherapist) or captured using for example Kinect cameras by thresholding the deviation of the ideal execution with the real measurement to find interesting points. Further indicators for important points in the data could be identified when the user aborts the interaction or communication with the system or the given task or other unforeseen interaction patterns occur.

Following this guideline the different requirements that were defined above are naturally met in the design of the tasks for the corpus. The involvement of the recorded subjects is asserted as they carry out a task that is chosen for their individual interests, for example to improve in a particular skill or in an aspect of their physical health. In many applications of the type described above, the repeatability of the task is naturally provided by their exercise nature. The personalization of the respective task can be naturally conducted by adjusting the difficulty of the subject matter in the teaching domain or the type and intensity of the exercises to monitor in a hospital situation.

A long term experiment could be in the first phase of a dataset to induce emotions and dispositions via specific and carefully selected stimuli and in the second task to focus on the ability of the system to detect this emotion/disposition and to regulate the interaction back to a positive state. This closed-loop experimental design would allow for the classification to detect the emotional state and afterwards to be able to influence the ongoing interaction with the means of positive feedback, help or supporting stimuli. However, a self rating should be included (see above) to have a manipulation check.

5 CONCLUSION

In this position paper, the common design process for affective data collections was reviewed and ideas were developed to categorize the necessary steps from choice of subjects over design of stimuli to the annotation of the material. In this work, we propose to shift the majority of work in the creative process from corpus design and recording towards annotation to achieve even more natural corpora. This goal is achieved by leveraging the intrinsic motivation of subjects in performing specific tasks that occur naturally or by design based on the subject group and then use human raters to annotate the engaging parts of the recording using hybrid labeling. Future implementations of the proposed paradigms should validate the feasibility of this approach.

ACKNOWLEDGEMENTS

This paper is based on work done within the Transregional Collaborative Research Centre SFB/TRR 62 *Companion-Technology for Cognitive Technical Systems* funded by the German Research Foundation (DFG). Markus Kächele is supported by a scholarship of the Landesgraduiertenförderung Baden-Württemberg at Ulm University.

REFERENCES

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., and Weiss, B. (2005). A database of german emotional speech. In *INTERSPEECH'05*, pages 1517–1520.

Codispoti, M., Surcinelli, P., and Baldaro, B. (2008). Watching emotional movies: Affective reactions and gender differences. *International Journal of Psychophysiology*, 69(2):90–95.

Daly, I., Malik, A., Hwang, F., Roesch, E., Weaver, J., Kirke, A., Williams, D., Miranda, E., and Nasuto, S. J. (2014). Neural correlates of emotional responses to music: An EEG study. *Neuroscience Letters*, 573(0):52–57.

Ekman, P., Sorenson, E. R., and Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164(3875):86–88.

Glodek, M., Reuter, S., Schels, M., Dietmayer, K., and Schwenker, F. (2013). Kalman filter based classifier fusion for affective state recognition. In *MCS*, volume 7872 of *LNCS*, pages 85–94.

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in Cognitive Sciences*, 16(9):458–466.

Healey, J. A. (2000). *Wearable and automotive systems for affect recognition from physiology*. PhD thesis, MIT.

Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., and Bartussek, D. (2005). A revised film set for the induction of basic emotions. *Cognition & Emotion*, 19(7):1095–1109.

Kächele, M., Glodek, M., Zharkov, D., Meudt, S., and Schwenker, F. (2014). Fusion of audio-visual features using hierarchical classifier systems for the recognition of affective states and the state of depression. In *Proceedings of ICPRAM*, pages 671–678. SciTePress.

Kächele, M., Schels, M., and Schwenker, F. (2014). Inferring depression and affect from application dependent meta knowledge. In *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, pages 41–48. ACM. Best entry for the Affect Recognition Sub-Challenge.

Kanade, T., Cohn, J., and Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000.*, pages 46–53.

Kreibig, S. D., Wilhelm, F. H., Roth, W. T., and Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear and sadness inducing films. *Psychophysiology*, 44(5):787–806.

Labouvie-Vief, G., Lumley, M. A., Jain, E., and Heinze, H. (2003). Age and gender differences in cardiac reactivity and subjective emotion responses to emotional autobiographical memories. *Emotion*, 3(2):115–126.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2005). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-6, University of Florida, Gainesville, FL, Gainesville, FL.

Lindquist, K. a., Siegel, E. H., Quigley, K. S., and Barrett, L. F. (2013). The hundred-year emotion war: are emotions natural kinds or psychological constructions? comment on Lench, Flores, and Bench (2011). *Psychological bulletin*, 139(1):255–63.

Nater, U. M., Abbruzzese, E., Krebs, M., and Ehlert, U. (2006). Sex differences in emotional and psychophysiological responses to musical stimuli. *Int J Psychophysiol*, 62(2):300–308.

Palm, G. and Glodek, M. (2013). Towards emotion recognition in human computer interaction. In *Neural Nets and Surroundings*, volume 19 of *Smart Innovation, Systems and Technologies*, pages 323–336.

Rösner, D., Frommer, J., Friesen, R., Haase, M., Lange, J., and Otto, M. (2012). LAST MINUTE: a multimodal corpus of speech-based user-companion interactions. In *Proc. of LREC*, pages 2559–2566.

Rukavina, S., Gruss, S., Tan, J.-W., Hrabal, D., Walter, S., Traue, H., and Jerg-Bretzke, L. (2013). The impact of gender and sexual hormones on automated psychobiological emotion classification. In *Proc of HCII*, volume 8008 of *LNCS*, pages 474–482.

Russell, J. A. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273 – 294.

Schels, M., Glodek, M., Palm, G., and Schwenker, F. (2013a). Revisiting AVEC 2011 — an information fusion architecture. In *Computational Intelligence in Emotional or Affective Systems*, Smart Innovation, Systems and Technologies, pages 385–393. Springer.

Schels, M., Kächele, M., Glodek, M., Hrabal, D., Walter, S., and Schwenker, F. (2014). Using unlabeled data to improve classification of emotional states in human computer interaction. *JMUI*, 8(1):5–16.

Schels, M., Scherer, S., Glodek, M., Kestler, H., Palm, G., and Schwenker, F. (2013b). On the discovery of events in EEG data utilizing information fusion. *Comput. Stat.*, 28(1):5–18.

Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social Science Information*, 44(4):695–729.

Schuller, B., Valstar, M., Eyben, F., McKeown, G., Cowie, R., and Pantic, M. (2011). AVEC 2011 — the first international audio visual emotion challenges. In *Proceedings of ACII (2011)*, volume 6975 of *LNCS*, pages 415–424. Part II.

Schwenker, F., Frey, M., Glodek, M., Kächele, M., Meudt, S., Schels, M., and Schmidt, M. (2014). A new multi-class fuzzy support vector machine algorithm. In *ANNPR*, volume 8774 of *LNCS*, pages 153–164.

Strauss, P.-M., Hoffmann, H., Minker, W., Neumann, H., Palm, G., Scherer, S., Traue, H., and Weidenbacher, U. (2008). The PIT corpus of german multi-party dialogues. In *Proc. of LREC*, pages 2442–2445.

Sun, X., Lichtenauer, J., Valstar, M., Nijholt, A., and Pantic, M. (2011). A multimodal database for mimicry analysis. In DMello, S., Graesser, A., Schuller, B., and Martin, J.-C., editors, *Proc of ACII*, volume 6974 of *LNCS*, pages 367–376.

Thiam, P., Meudt, S., Kächele, M., Palm, G., and Schwenker, F. (2014). Detection of emotional events utilizing support vector methods in an active learning hci scenario. In *Proc. of Emotion Representations and Modelling for HCI Systems*, ERM4HCI '14, pages 31–36. ACM.

Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., Schnieder, S., Cowie, R., and Pantic, M. (2013). Avec 2013: The continuous audio/visual emotion and depression recognition challenge. In *Proceedings of AVEC 2013*, pages 3–10. ACM.

Walter, S., Scherer, S., Schels, M., Glodek, M., Hrabal, D., Schmidt, M., Böck, R., Limbrecht, K., Traue, H., and Schwenker, F. (2011). Multimodal emotion classification in naturalistic user behavior. In Jacko, J., editor, *Proc. of HCII*, volume 6763 of *LNCS*, pages 603–611.

Walter, S., Wendt, C., Bhnke, J., Crawour, S., Tan, J., Chan, A., Limbrecht, K., Gruss, S., and Traue, H. (2013). Similarities and differences of emotions in human-machine and human-human interaction: what kind of emotions are relevant for future companion systems. *Ergonomics*, 57(3):374–386.

Wöllmer, M., Kaiser, M., Eyben, F., Schuller, B., and Rigoll, G. (2013). LSTM-modeling of continuous emotions in an audiovisual affect recognition framework. *Image and Vision Computing*, 31(2):153 – 163.