NOVA: A Tool for Explanatory Multimodal Behavior Analysis and Its Application to Psychotherapy

Tobias Baur[1(✉)], Sina Clausen[2], Alexander Heimerl[1], Florian Lingenfelser[1], Wolfgang Lutz[2], and Elisabeth André[1]

[1] Lab for Human-Centered Multimedia, Augsburg University, Augsburg, Germany
{baur,heimerl,lingenfelser,andre}@hcm-lab.de
[2] Klinische Psychologie und Psychotherapie, Trier University, Trier, Germany
{clausen,lutz}@uni-trier.de

Abstract. In this paper, we explore the benefits of our next-generation annotation and analysis tool NOVA in the domain of psychotherapy. The NOVA tool has been developed, tested and applied in behaviour studies for several years and psychotherapy sessions offer a great way to expand areas of application into a challenging yet promising field. In such scenarios, interactions with patients are often rated by questionnaires and the therapist's subjective rating, yet a qualitative analysis of the patient's non-verbal behaviours can only be performed in a limited way as this is very expensive and time-consuming. A main aspect of NOVA is the possibility of applying semi-supervised active learning where Machine Learning techniques are already used during the annotation process by giving the possibility to pre-label data automatically. Furthermore, NOVA provides therapists with a confidence value of the automatically predicted annotations. This way, also non-ML experts get to understand whether they can trust their ML models for the problem at hand.

Keywords: Annotation tools · Psychotherapy · Cooperative Machine Learning · Explainable AI

1 Introduction

In psychotherapy emotions appear to have a central role for the activation of change processes and therapeutic success [25,32]. Most mental disorders are characterised by disturbed affect (e.g. depression) and dysfunctional emotion regulation (e.g. substance abuse disorders) [17,22]. Many therapeutic approaches (e.g. emotion-focused therapy, schema therapy, dialectical-behavioural therapy) emphasise the need of changing emotional coping styles during the therapy process [32]. Multiple studies have shown that psychotherapy is generally effective

This work has received funding from the BMBF under FKZ 01IS17074, FMLA, the DFG under project number 392401413, DEEP and LU 660/10-1, LU660/8-1.

[8]. Yet, there is often high variability in outcomes [9]. Patient-focused psychotherapy research deals with the question how, when and why psychotherapy works in order to improve therapeutic success for the individual patient [20]. This field of research makes use of frequently repeated measures, mainly questionnaires, to examine session processes and outcomes (routine outcome monitoring, feedback) [18,21]. It includes for example the assessment of current symptoms and common factors (e.g. therapeutic relationship, problem solving, problem actuation) to predict symptom change (e.g. sudden gains and losses) and provide empirically supported feedback to therapists [19,21]. In this context, additional or alternative data sources would be highly valuable because humans, especially patients with depressive symptoms, tend to have a distorted recall of emotions and mood [3]. Further, the reported subjective feeling captures one component of emotions. Complementary information about the more objective (bodily) expression is of major importance to better understand interpersonal processes and communication (e.g. [15]).

Thus, the analysis of emotions by means of nonverbal signals in recorded psychotherapy sessions bears great potential for advancing both, the psychotherapy process as well as outcome research. In order to analyse recorded therapy sessions for identifying relevant behaviour patterns with this degree of refinement, in classical approaches a huge bottleneck is given by the necessity of manual labelling effort. That means, segments in the observed material need to be labelled using sets of discrete classes or continuous scores, e. g. a certain type of gesture, a social situation, or the emotional state of a person. Especially in the field of practical psychotherapy, where thousands of hours of data are generated, manually annotating would be an overambitious task. A solution to this problem is exploitation of computational power to accomplish some of the annotation work automatically. However, to ensure the quality of the predicted annotations this still requires human supervision to identify and correct errors. To keep human effort as low as possible, it is useful to understand why a model makes wrong assumptions. Therefore, it is not only important to provide tools that ease the use of semi-automated labelling, but also to increase the transparency of the decision process. By visualising the predictions, therapists get an idea about the strengths and weaknesses of the underlying classification model and can immediately decide which parts of a prediction are worth keeping. Ideally, the system even guides the users' attention towards parts where manual revision is necessary. Once an annotation has been revised, the model can be retrained by end-users of the system themselves to improve its performance for the next cycle. This procedure can be repeated until a desired performance is reached.

Once the model achieves a satisfying performance it may be used to predict new, unlabelled sessions in a fully automated manner. This automated analysis of emotional expression during the session can provide vital additional information that helps to understand the emotional activation of the patient and the emotional interaction with the therapist (e.g. affective co-regulation, synchrony). In this paper, we first introduce a next-generation annotation tool called NOVA, which implements a cooperative machine learning workflow. In partic-

ular, NOVA offers semi-automated annotations and provides visual feedback to inspect and correct machine-generated labels. We further report on a pilot study that describes first experiences with NOVA in the practical application in the area of psychotherapy.

2 Related Work

2.1 Annotation Tools

In the past, several annotation tools with focus on analysing social signals in various contexts have been developed. The general user interface of NOVA has been inspired by existing annotation tools. Prominent examples include ELAN [33], ANVIL [16], and EXMARALDA [28]. These tools offer layer-based tiers to insert time-anchored labelled segments, that we call *discrete* annotations. *Continuous* annotations, on the other hand, allow an observer to track the content of an audiovisual stimulus over time based on a continuous scale. One of the first tools that allow annotators to trace emotional content in real-time on two dimensions (activation and evaluation) was FEELTRACE [6]. Its descendant GTRACE (general trace) [7] allows the user to define their own dimensions and scales. More recent tools to accomplish continuous descriptions are CARMA (continuous affect rating and media annotation) [10] and DARMA (dual axis rating and media annotation) [11].

Unfortunately, almost all of the tools offer none or only little automation. In former studies, labelling of several hours of interaction turned out to be an extremely time consuming task, so methods to automate the coding process were highly desirable. In the targeted scenario, bearing the potential of several thousands of hours of recorded sessions, automation of the labelling process becomes imperative.

2.2 Active and Cooperative Machine Learning

A common approach to reduce human labelling effort is the selection of instances for manual annotation based on active learning techniques. The basic idea is to forward only instances with low prediction certainty or high expected error reduction to human annotators [29]. Estimation of most informative instances is an art of its own. A whole range of options to choose from exist, such as calculation of 'meaningful' confidence measures, detecting novelty (e. g. by training auto-encoders and seeing for the deviation of input and output when new data runs through the auto-encoder), estimating the degree of model change the data instance would cause (e.g. seeing whether knowing the label of a data point would make a change to the model at all), or trying to track 'scarce' instances, e.g. trying to find those data instances that are rare in terms of the expected label. Further, more sophisticated approaches aggregate the results of machine learning and crowd-sourcing processes to increase the efficiency of the labelling process. Kamar et al. [14] make use of learned probabilistic models to fuse results

from computational agents and human annotators. They show how to allocate tasks to coders in order to optimise crowdsourcing processes based on expected utility. Relatively little attention has been paid, however, to the question of how to make these techniques available to human annotators. There is a high demand for annotation tools that integrate cooperative machine learning in order to reduce human effort.

Most studies in this area focus on the gain obtained by the application of specific active learning techniques. However, little emphasis is given to the question of how to assist users in the application of these techniques for the creation of their own corpora. While the benefits of integrating active learning with annotation tasks has been demonstrated in a variety of experiments, annotation tools that provide users with access to active learning techniques are rare.

3 NOVA Tool

The NOVA tool aims to enhance the standard annotation process with the latest developments from contemporary research fields such as Cooperative Machine Learning by giving annotators easy access to automated model training and prediction functionalities, as well as sophisticated explanation algorithms via its user interface.

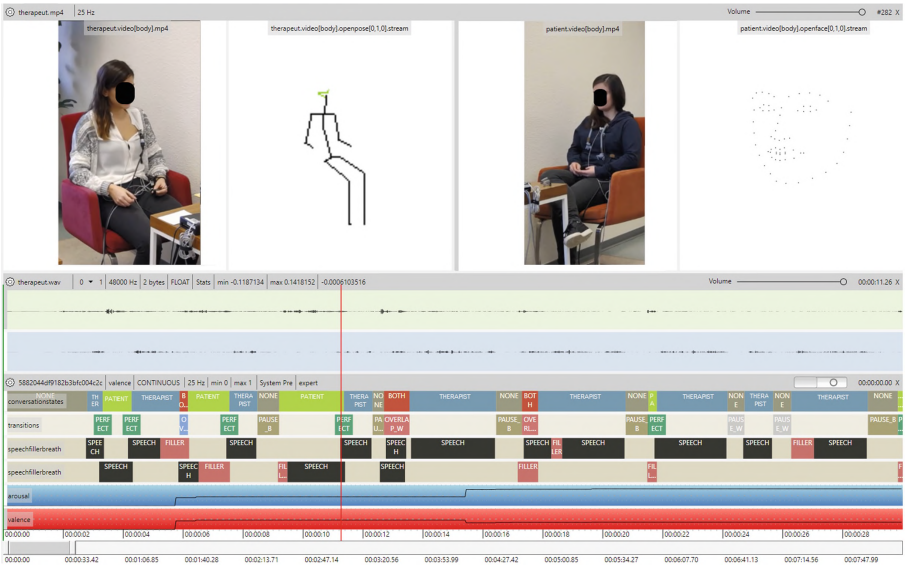

Fig. 1. NOVA allows to visualise various media and signal types and supports different annotation schemes. This figure illustrates an instance of a clinical therapy session with a therapist and a patient. From top downwards: full-body videos along with skeleton and face tracking, and audio streams of two persons during an interaction. In the lower part, several discrete and continuous annotation tiers are displayed.

The NOVA user interface has been designed with a special focus on the annotation of long and continuous recordings involving multiple modalities and subjects. A screenshot of a loaded clinical therapy session is shown in Fig. 1. On the top, several media tracks are visualised and ready for playback. Note that the number of tracks that can be displayed at the same time is not limited and various types of signals (video, audio, facial features, skeleton, depth images, etc.) are supported. In the lower part, we see multiple annotation tracks describing the visualised content with multiple types of annotations (Discrete, Free and Continuous).

To support a collaborative annotation process, NOVA maintains a database back-end, which allows users to load and save annotations from and to a MongoDB[1] database running on a central server. This gives annotators the possibility to immediately commit changes and follow the annotation progress of others. Beside human annotators, a database may also be visited by one or more "machine users". Just like a human operator, they can create and access annotations. Hence, the database also functions as a mediator between human and machine. NOVA provides instruments to create and populate a database from scratch. At any time new annotators, schemes and additional sessions can be added. NOVA provides several functions to process the annotations created by multiple human or machine annotators. For instance, statistical measures such as Cronbach's α, Pearson's correlation coefficient, Spearman's correlation coefficient or Cohen's κ can be applied to identify inter-rater agreement. Thus, the foundations have been laid to fine-tune the number of annotators based on inter-rater agreement in order to further reduce work load by allocating human resources to instances that are difficult to label (see [34]).

Tasks related to machine learning (ML) are handed over and executed by our open-source Social Signal Interpretation (SSI) framework [31]. Since SSI is primarily designed to build online recognition systems, a trained model can be directly used to detect social cues in real-time. A typical ML pipeline starts by prepossessing data to input data for the learning algorithm, a step known as *feature extraction*. An XML template structure is used to define extraction chains from individual SSI components. A dialogue helps users to extract features by selecting an input stream and a number of sessions. The result of the operation is stored as a new signal in the database. This way, feature streams can be reviewed in NOVA and accessed by all users. Based on the extracted features, a classifier, which may also be added using XML templates, can be trained. Alternatively, NOVA supports Deep and Transfer Learning by providing Python interfaces to Tensorflow and Keras. This way convolutional networks may be trained, respectively retrained, based on annotations saved in NOVA's annotation database on raw video data.

[1] https://www.mongodb.com/.

4 Cooperative Machine Learning

In this paper, we subsume learning approaches that efficiently combine human intelligence with the machine's ability of rapid computation under the term *Cooperative Machine Learning* (CML). In Fig. 2, we illustrate our approach to CML, which creates a loop between a machine learned model and human annotators: an initial model is trained (1) and used to predict unseen data (2). An active learning module then decides which parts of the prediction are subject to manual revision by human annotators (3 + 4). Afterwards, the initial model is retrained using the revised data (5). Now the procedure is repeated until all data is annotated. By actively incorporating the user into the loop it becomes possible to interactively guide and improve the automatic predictions while simultaneously obtaining an intuition for the functionality of the classifier.

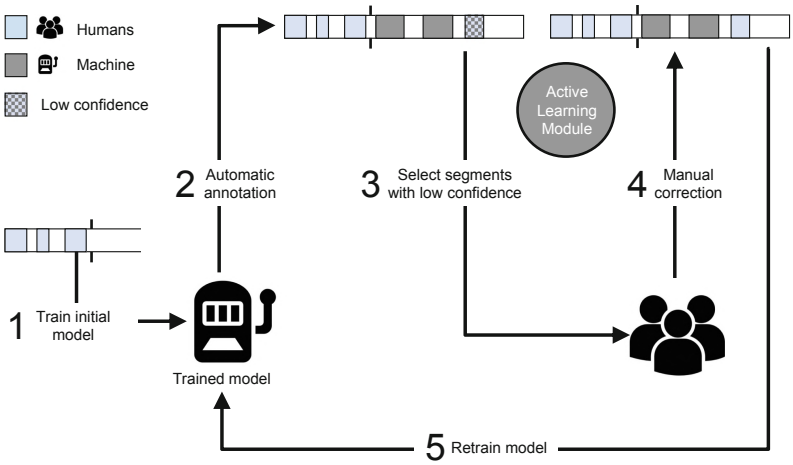

Fig. 2. The scheme depicts the general idea behind Cooperative Machine Learning (CML): (1) An initial model is trained on partially labelled data. (2) The initial model is used to automatically predict unseen data. (3) Labels with a low confidence are selected and (4) manually revised. (5) The initial model is retrained with the revised data.

However, the approach does not only bear the potential to considerably cut down manual efforts, but also to come up with a better understanding of the capabilities of the classification system. For instance, the system may quickly learn to label some simple behaviours, which already facilitates the work load for human annotators at an early stage. Then, over time, it could learn to cope with more complex social signals as well, until at some point it is able to finish the task in a completely automatic manner.

To automatically finish an annotation, the user either selects a previously trained model or temporarily builds one using the labels on the current tier.

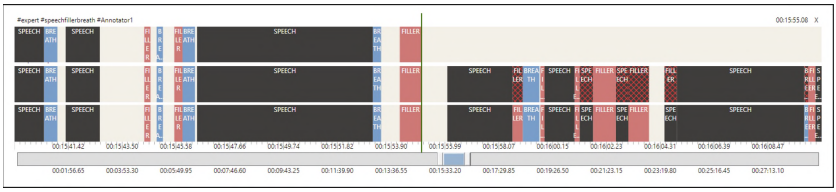

Fig. 3. The upper tier shows a partly finished annotation. ML is now used to predict the remaining part of the tier (middle), where segments with a low confidence are highlighted with a red pattern. The lower tier shows the final annotation after manual revision. (Color figure online)

An example before and after the completion is shown in Fig. 3. Note that labels with a low confidence are highlighted with a pattern. This way, the annotator can immediately see how well the prediction worked. To evaluate the efficiency of the integrated CML strategy, in our earlier work [30] we performed a simulation study on an audio-related labelling task. Following this approach we were able to reduce the initial annotation labour of 9.4 h to 5.9 h, which is a reduction of 37.23%.

5 Pilot Study in Psychotherapy Research

The application of NOVA for the analysis of psychotherapy sessions could lead to multiple advantages for understanding nonverbal signals and communication in psychotherapy. Yet, the transfer to naturalistic conditions sometimes involves obstacles. Therefore, we conducted a pilot study in order to evaluate the applicability of NOVA for this setting.

5.1 Study Setup

The pilot study comprises therapy videos from a protocol-based treatment for test anxiety [26]. The protocol includes cognitive-behavioural therapy components and imagery rescripting which is a experiential and emotionally activating technique. The sample includes therapy videos from 12 patient-therapist dyads with 6 sessions each with a duration of 50–60 min. The patients were students with high scores on test anxiety (partly with depressive symptoms). The therapists were three graduates of a masters' programme of psychology who were awaiting the beginning of their clinical training and two PhD-students in clinical psychology who each had prior experience of one year as a clinician.

The starting module involved the installation and database set up at the outpatient clinic as well as an introduction for the main features of NOVA. Since the therapy videos are very confidential material, all annotations need to be conducted within the clinic. First, the recorded therapy videos were converted to suit the requirements of NOVA (i.a. single video streams for each role). The data structure of sessions with different roles (therapist and patient) which can

be loaded simultaneously or separately is ideal for reducing input during the annotation process and yet being able to visually observe interpersonal dynamics. In the following, body features from OpenPose [5] and facial landmarks from OpenFace [2] were extracted. Concerning the performance of the feature extraction, the Openface and Openpose tracking algorithms were able to successfully extract features in about 70% of the test videos. In the remaining videos the feature extraction partly failed due to known challenges, such as problematic head and body orientations (facing/sitting non-frontal to the camera) or characteristics of the person in the video (beard, hat etc.).

For the manual annotations two advanced master students awaiting the beginning of their clinical training were trained in NOVA and the circumplex-model of emotions (valence, arousal) in a 3h-training session. A training video was annotated and discussed in order to achieve a common understanding. A difficulty in this context evolved through cases with incongruent signals on the verbal and non-verbal canal or suppressed emotions. It was decided to annotate by means of the global impression and to give the nonverbal signals a higher weight in unclear situations. This is in line with findings of a higher importance of nonverbal signals in unclear communicative situations (nonverbal dominance, e.g. [13]). Further, NOVA was introduced to six additional students in the master programme of psychology in the context of a seminar. They rated each about 6 h of recorded sessions and also gave qualitative feedback. Since the rating is still in progress, this paper presents qualitative feedback of the experiences with applying NOVA so far.

5.2 Results

The qualitative feedback from the eight raters (two with >20 h of rating experience with NOVA and six with about 6 h) included the following aspects. Every rater gave positive feedback concerning the use of NOVA being "intuitive" and ergonomic in main functions (e.g. start/ stop, Live-Mode, the possibility to correct and overwrite incorrect annotations, adaptation of the pace of the annotation stream). Critical feedback was mainly concerned with the general task of continuously annotating arousal and valence in recorded sessions of psychotherapy. Raters reported that, despite psychological knowledge and training, the external rating of emotions with valence and arousal is still prone to subjective interpretations. This was reflected in low inter rater agreement in some segments, especially for persons with low congruence of verbal and nonverbal signals or low expressiveness. Here, NOVA was helpful in identifying the corresponding parts of the interaction. Thus, a good training and the creation of merged annotations from multiple ratings seems important to achieve more objective training data. The rating is challenging because the context often creates ambiguity. For example laughter implies positive valence yet the arousal can be different. In general, laughter often marks joy and therefore a higher arousal. Yet, in a context of heightened tension and personal distress it can be a sign of relief and thus a lower arousal. The context-dependent ambiguity possibly leads to lower quality of the training data and basis for the machine learning process. Further, manual

continuous annotation for longer videos was described to be "very tiring", thus the required annotated video time should be as short as possible. Some raters said that it is especially difficult to rate the same video twice (once for each valence- and arousal dimension). A simultaneous 2-dimensional rating (e.g. by means of a joystick) might help to reduce to rating time, yet might lead to other problems, like a drop in annotation accuracy due to multi-tasking on the side of the annotators. Additional aspects of qualitative feedback for functions in NOVA apart from the direct rating process involved the following positive points: The possibility to conduct all relevant steps of nonverbal behaviour analysis within one tool (annotation, model training, model evaluation) leads to high usability. Multiple annotators can be assigned to databases and annotations with different rights, this is very practical. NOVA offers a high level of adaptability and flexibility concerning the concrete steps (e.g. which annotations to merge, which models to train and apply). The study participants pointed out that the rating effort to build a first model appears quite high, especially for continuous variables in naturalistic settings where strong emotional changes do not occur very often. Thus, for the future the use of pre-trained models under similar conditions would seem helpful. This is likely feasible as NOVA also allows for multi-database training. In summary, the experiences with NOVA so far indicate that it encompasses many important functions for the analysis of emotions in psychotherapy sessions.

5.3 Perspectives for NOVA in Psychotherapy Research and Practice

Multiple topics of major importance in psychotherapy research can presumably profit from a successful application of NOVA, both in process and outcome research. In process research an automated emotion recognition can add knowledge about beneficial therapeutic processes (e.g. successfully targeting emotions, characteristics of sessions prior to sudden gains or early positive change of symptoms, associations with the therapeutic relationship) [1,27]. Further, data concerning the emotions could be linked to problematic developments (e.g. alliance ruptures, sudden losses, dropout) [23]. After having achieved knowledge about the associations between emotional activation and co-regulation patterns with developments in therapy it can be included into the routine feedback systems (e.g. [21]). Feedback could include for example the following "Your patient had almost no phases of positive valence in the last session, you may consider to improve resource activation and supporting needs of the patient". For outcome research emotional activation could potentially be employed as an outcome measure. One could expect for example for depressed patients an increase in emotional variability and positive emotions in the course of therapy [4]. Emotional activation in the beginning compared to the end of therapy could act as a measure of change. Further, patients would possibly reveal atypical patterns of emotional co-regulation (i.e. synchrony) which could change towards more typical patterns during the course of therapy and thus act as another indicator of change [12,24].

For psychotherapists in clinical practice NOVA could be a helpful tool to achieve more objective insight to in-session processes. For example, the therapist could use it to reveal discrepancies between self-perception of the patient and external perception and discuss it with the patient. A potential scenario would be, that a patient is convinced that his grief is clearly observable for others and, in contrast, there is no sign of negative valence during the sessions. The therapist could also use it for self-reflection. For example if he has very low arousal and negative valence with one patient compared to others he could use it as an indicator to make use of supervision.

6 Conclusion

In this paper we presented an initial application of the NOVA annotation and analysis tool in the context of patient-focused psychotherapy. NOVA offers a collaborative workflow for multiple types of annotation tasks. Additionally it provides interfaces to machine-learning techniques that allow even non-experts to make use of these technologies in order to speed up the annotation labour.

The qualitative feedback for working with NOVA was consistently very positive. The experiences with feature extraction, annotation and machine learning in NOVA under real-life conditions give reason for optimism concerning the further application of NOVA in psychotherapy research and practice.

With little effort, models can be retrained and fine-tuned to specific scenarios and types of patients without the need of programming-skills. Yet, more technically interested users can also extend NOVA's ML tools by adding new templates. This way NOVA is not limited to current state of the art methods such as Deep Neural Networks but is also extendable in the future. Additionally, we are working on extending NOVA to provide capabilities for the latest explainable AI techniques on pre-trained as well as self-trained models. This way, users should get an even better understanding when they can trust their ML model and what might cause issues, respectively when more training examples are required.

NOVA is open-source software and available on Github: https://github.com/hcmlab/nova.

References

1. Atzil-Slonim, D., et al.: Emotional congruence between clients and therapists and its effect on treatment outcome. J. Couns. Psychol. **65**(1), 51–64 (2018)
2. Baltrušaitis, T., Robinson, P., Morency, L.P.: OpenFace: an open source facial behavior analysis toolkit. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–10, March 2016. https://doi.org/10.1109/WACV.2016.7477553
3. Ben-Zeev, D., Young, M.A., Madsen, J.W.: Retrospective recall of affect in clinically depressed individuals and controls. Cogn. Emot. **23**(5), 1021–1040 (2009)
4. Bylsma, L.M., Morris, B.H., Rottenberg, J.: A meta-analysis of emotional reactivity in major depressive disorder. Clin. Psychol. Rev. **28**(4), 676–691 (2008)

5. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y.: OpenPose: realtime multi-person 2D pose estimation using part affinity fields. arXiv preprint arXiv: 1812.08008 (2018)
6. Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schröder, M.: 'FEELTRACE': an instrument for recording perceived emotion in real time. In: ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion (2000)
7. Cowie, R., McKeown, G., Douglas-Cowie, E.: Tracing emotion: an overview. IJSE **3**(1), 1–17 (2012)
8. Cuijpers, P., van Straten, A., Andersson, G., van Oppen, P.: Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J. Consult. Clin. Psychol. **76**(6), 909–922 (2008)
9. Delgadillo, J., Moreea, O., Lutz, W.: Different people respond differently to therapy: a demonstration using patient profiling and risk stratification. Behav. Res. Ther. **79**, 15–22 (2016)
10. Girard, J.M.: CARMA: software for continuous affect rating and media annotation. J. Open. Res. Softw. **2**(1), e5 (2014)
11. Girard, J.M., Wright, A.G.C.: DARMA: dual axis rating and media annotation (2016)
12. Hofmann, S.G.: Interpersonal emotion regulation model of mood and anxiety disorders. Cogn. Ther. Res. **38**(5), 483–492 (2014)
13. Jacob, H., Kreifelts, B., Brück, C., Nizielski, S., Schütz, A., Wildgruber, D.: Nonverbal signals speak up: association between perceptual nonverbal dominance and emotional intelligence. Cogn. Emot. **27**(5), 783–799 (2013)
14. Kamar, E., Hacker, S., Horvitz, E.: Combining human and machine intelligence in large-scale crowdsourcing. In: van der Hoek, W., Padgham, L., Conitzer, V., Winikoff, M. (eds.) International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012, Valencia, Spain, 4–8 June 2012, vol. 3 , pp. 467–474. IFAAMAS (2012)
15. Kennedy-Moore, E., Watson, J.C.: How and when does emotional expression help? Rev. Gen. Psychol. **5**(3), 187–212 (2001)
16. Kipp, M.: ANVIL: the video annotation research tool. In: Handbook of Corpus Phonology. Oxford University Press (2013)
17. Lukas, C.A., Ebert, D.D., Fuentes, H.T., Caspar, F., Berking, M.: Deficits in general emotion regulation skills-evidence of a transdiagnostic factor. J. Clin. Psychol. **74**(6), 1017–1033 (2018)
18. Lutz, W., et al.: Chancen von e-mental-health und eprozessdiagnostik in der ambulanten psychotherapie: Der trierer therapie navigator. Verhaltenstherapie 1–10 (2019)
19. Lutz, W., et al.: The ups and downs of psychotherapy: sudden gains and sudden losses identified with session reports. Psychother. Res. **23**(1), 14–24 (2013)
20. Lutz, W., de Jong, K., Rubel, J.: Patient-focused and feedback research in psychotherapy: where are we and where do we want to go? Psychother. Res. **25**(6), 625–632 (2015)
21. Lutz, W., Rubel, J.A., Schwartz, B., Schilling, V., Deisenhofer, A.K.: Towards integrating personalized feedback research into clinical practice: development of the Trier Treatment Navigator (TTN). Behav. Res. Ther. **120**, 103438 (2019)
22. Marwood, L., Wise, T., Perkins, A.M., Cleare, A.J.: Meta-analyses of the neural mechanisms and predictors of response to psychotherapy in depression and anxiety. Neurosci. Biobehav. Rev. **95**, 61–72 (2018)

23. Paulick, J., et al.: Nonverbal synchrony: a new approach to better understand psychotherapeutic processes and drop-out. J. Psychother. Integr. **28**(3), 367–384 (2018)
24. Paulick, J., et al.: Diagnostic features of nonverbal synchrony in psychotherapy: comparing depression and anxiety. Cogn. Ther. Res. **42**(5), 539–551 (2018)
25. Peluso, P.R., Freund, R.R.: Therapist and client emotional expression and psychotherapy outcomes: a meta-analysis. Psychotherapy **55**(4), 461–472 (2018). (Chicago, Ill.)
26. Prinz, J.N., Bar-Kalifa, E., Rafaeli, E., Sened, H., Lutz, W.: Imagery-based treatment for test anxiety: a multiple-baseline open trial. J. Affect. Disord. **244**, 187–195 (2019)
27. Rubel, J.A., Rosenbaum, D., Lutz, W.: Patients' in-session experiences and symptom change: session-to-session effects on a within- and between-patient level. Behav. Res. Ther. **90**, 58–66 (2017)
28. Schmidt, T.: Transcribing and annotating spoken language with EXMARaLDA. In: 2004 Proceedings of the International Conference on Language Resources and Evaluation: Workshop on XML Based Richly Annotated Corpora, Lisbon, pp. 879–896. ELRA (2004). eN
29. Settles, B.: Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, San Rafael (2012)
30. Wagner, J., Baur, T., Zhang, Y., Valstar, M.F., Schuller, B., André, E.: Applying cooperative machine learning to speed up the annotation of social signals in large multi-modal corpora. arXiv preprint arXiv:1802.02565 (2018)
31. Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., André, E.: The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time. In: Proceedings of the 21st ACM International Conference on Multimedia, pp. 831–834. ACM (2013)
32. Whelton, W.J.: Emotional processes in psychotherapy: evidence across therapeutic modalities. Clin. Psychol. Psychother. **11**(1), 58–71 (2004)
33. Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H.: ELAN: a professional framework for multimodality research. In: Calzolari, N., et al. (eds.) Proceedings of the Fifth International Conference on Language Resources and Evaluation, LREC 2006, Genoa, Italy, 22–28 May 2006, pp. 1556–1559. European Language Resources Association (ELRA) (2006)
34. Zhang, Y., Michi, A., Wagner, J., André, E., Schuller, B., Weninger, F.: A generic human-machine annotation framework based on dynamic cooperative learning. IEEE Trans. Cybern. 1–10 (2019)