UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

NovA: automated analysis of nonverbal signals in social interactions

Tobias Baur, Ionut Damian, Florian Lingenfelser, Johannes Wagner, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Baur, Tobias, Ionut Damian, Florian Lingenfelser, Johannes Wagner, and Elisabeth André. 2013. "NovA: automated analysis of nonverbal signals in social interactions." In *Human behavior understanding: 4th International Workshop, HBU 2013, Barcelona, Spain, October 22, 2013*, edited by Albert Ali Salah, Hayley Hung, Oya Aran, and Hatice Gunes, 160–71. Berlin [u.a.]: Springer. https://doi.org/10.1007/978-3-319-02714-2_14.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
Deutsches Urheberrecht
Weitere Informationen finden Sie unter: / For more information see:
https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/

NovA: Automated Analysis of Nonverbal Signals in Social Interactions

Tobias Baur, Ionut Damian, Florian Lingenfelser, Johannes Wagner, and Elisabeth André

Human Centered Multimedia, Augsburg University,
Universitätsstr. 6a, 86159 Augsburg, Germany
{baur,damian,lingenfelser,wagner,andre}@hcm-lab.de
http://www.hcm-lab.de

Abstract. Previous studies have shown that the success of interpersonal interaction depends not only on the contents we communicate explicitly, but also on the social signals that are conveyed implicitly. In this paper, we present NovA (NOnVerbal behavior Analyzer), a system that analyzes and facilitates the interpretation of social signals conveyed by gestures, facial expressions and others automatically as a basis for computer-enhanced social coaching. NovA records data of human interactions, automatically detects relevant behavioral cues as a measurement for the quality of an interaction and creates descriptive statistics for the recorded data. This enables us to give a user online generated feedback on strengths and weaknesses concerning his social behavior, as well as elaborate tools for offline analysis and annotation.

1 Introduction

In a conversation between humans, information is not only shared in an explicit manner. On the contrary, a myriad of implicit social signals is communicated that may even have a deeper influence on the outcome of a conversation than the word meanings themselves. There is empirical evidence that our impact on others is shaped less by the contents of an utterance, but rather by the accompanying social signals. According to studies by Albert Mehrabian [1], the contents of an utterance only contribute with 7% to its success while the impact of the vocal signals (conveyed by the nuances of voice) and the non-verbal signals (including body language and facial expressions) is much higher with 38% and 55% respectively.

In our research, we present NovA (NOnVerbal behavior Analyzer), a system that analyzes and interprets these signals automatically as a starting point for social coaching in human-human settings. NovA does not only allow us to record data of social interactions in a systematic manner, but it also enables the learners to inspect previous interactions and provides them with an objective report of the social interactions.

Automated behavior analysis is not only of benefit in the context of social training. It may also support researchers of multidisciplinary areas in their daily

work. Scientific disciplines, such as psychology, ethnology, anthropology and others, have been concerned with the systematic exploration of human behavior for a long time. Thereby, an important component of a regular work flow is the annotation of audiovisual recordings of human behavior. The annotation of such recordings requires several iterations and is very time-consuming. With our analysis tool NovA, we aim to accomplish most of these iterations in a fully automated manner by creating a variety of diagrams including bar charts, heat maps, timeline diagrams with automated labeling that help to point out characteristics encountered in interpersonal interaction.

2 Related Work

To get insight into human social behavior, researchers have to rely on a large variety of information, e.g. video recordings of face-to-face meetings and acceleration sensor-based data of the user's daily activities [2]. Progress in the field has been boosted by a variety of annotation tools that facilitate the labeling of corpora at different levels of granularity following a pre-defined coding scheme. Examples include Elan, [3] Anvil [4] and Exmeralda [5]. However, since the manual annotation of data is rather time-consuming, methods to automate the coding process are highly desirable.

Techniques for the automated analysis of social behaviors patterns were pioneered by Pentland and his group at MIT Media Lab with the development of wearable devices, so-called sociometers, to capture the people's verbal and non-verbal signals. They investigated not only the social behaviors of people engaged in face-to-face conversations [6], but also analyzed interaction patterns from larger groups of people using smartphones with dedicated sensors [7]. Social constructs that have been investigated by Pentland and others include internal user states, such as interest [8], engagement [9] and emotions [10], and personality traits [11], such as dominance and extraversion.

To analyze social behaviors, a large variety of verbal and non-verbal cues has been taken into account. Dong and colleagues [12] analyze speech activity and fidgeting, i.e. the amount of movement in a person's hands and body, to detect functional roles in a group. Hung and Gatica-Perez [13] studied audio cues (such as overlapping speech), video cues (such as motion energy), and audiovisual cues (such as the amount of movement during speech) to determine the level of group cohesion in meetings. Methods have been developed to detect a user's emotions from various modalities including facial expressions [14], gestures [15], speech [16], postures [17] and physiological measurements [18]. Also, multimodal approaches to improve emotion recognition accuracy are reported, mostly by exploiting audiovisual combinations [19] [20]. Results suggest that integrated information from audio and video leads to improved classification reliability compared to a single modality - even with fairly simple fusion methods.

NovA combines work on annotation tools with technologies to automatically analyze human behavior. The user interface of NovA has been inspired by existing annotation tools and makes use of multiple tracks to code relevant social features. However, unlike conventional annotation tools, NovA performs the

segmentation and labeling of the data completely automatically. Most tools for automated behavior analysis rely on an offline classification of the recorded data, where relevant features are extracted from the data and mapped onto given social constructs. A problem with this approach results from the subjectivity of data interpretation. NovA distinguishes from earlier work on automated behavior analysis by combining a descriptive with interpretative coding approaches. Not only does it present the user with an interpretation of the data in terms of higher-level social constructs, but also visualizes statistics of the descriptive features on which the interpretation of the data is based. In addition, it is able to provide verbalized explanations. The motivation behind this approach is to increase the transparency of the system for the social coaches and their trainees.

3 Foundations of the NovA Approach

Compared to natural language, the capabilities of social cues to convey meaning are strongly limited. Their strength lies, however, in the communication of implicit information, such as the engagement of people in a conversation and the self efficacy they convey when presenting themselves. Based on interviews with job counselors we decided to focus on four user states that are implicitly conveyed in interviews by body postures and gestures *Social Attraction*, *Engagement*, *Self Efficacy* and *Attitude*.

Social Attraction refers to the amount of appreciation a person evokes in others [21]. The relation of body language and social attraction has been investigated in various studies in social sciences. McGinley et al. [22] conclude that open body positions are usually received as more positive than positions with arms or legs crossed. According to Schouwstra and Hoogstraten [23] upright postures with the head up receive more positive judgments than the opposite.

Sidner and colleagues define [24] *Engagement* as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake". Pease [25] demonstrates how engagement is portrayed by an orientation of the body and face towards the interlocutor. Another aspect of engagement is the overall amount of movements. Here it is necessary to distinguish whether the user is speaking or listening. While speakers tend to show their engagement using a high amount of overall activity, in the role of a listener, interlocutors should show less overall activity because such a behavior is usually interpreted as a sign of distraction.

People with a high amount of *Self Efficacy* are "confident that they will be able to master difficult situations" [26]. Self efficacy is usually conveyed by calm, fluid and high energy movements while quick and jerky movements tend to make a person appear nervous. In addition, a high amount of self manipulations, such as scratching one's head, reveals the anxiety of people in a social situation. Pease [25] provides various examples of body postures that signal a high amount of self efficacy, such as placing both feet apart or both hands behind the head with the elbows facing outward.

In psychology, the term *Attitude* refers to the expression of favor or disfavor towards a particular person or theme [27]. Usually, open body postures, such as

opened arms, are interpreted as a sign of willingness to cooperate while closed body postures, such as crossing one's arms, rather communicate the opposite [25].

4 Social Cue Recognition

For recording and preprocessing human behavior data, NovA relies on our previously developed Social Signal Interpretation framework (SSI) [28] which supports both frame-by-frame and event-based annotations. In the case of frame-by-frame annotations, a value referring to a particular attribute, such as the distance between the two hands of a person, is computed at each point in time. For event-based annotations, we implemented a mechanism that triggers an event each time the beginning or the end of a social cue, such as a particular arm configuration, is detected.

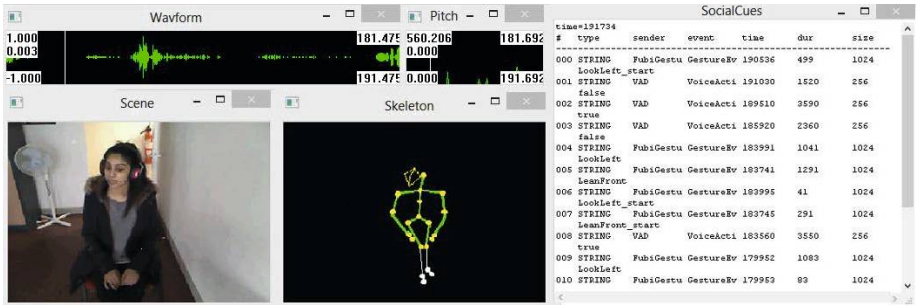

Fig. 1. The recording process with skeleton and face tracking, audio graph, audio pitch information, as well as the event board that shows detected social cues, such as gestures, head poses, voice activity detection etc

The detected events are saved in an XML-based structure including a synchronized timestamp and the event's duration. Figure 1 shows the interface for the recording of data. It illustrates the Kinect skeleton and facial tracking, as well as online recognized social cues, waveform graph and audio pitch. In the following subsections we present social cues that are recognizable by NovA.

4.1 Gesture and Posture Detection

For event-based gesture analysis, NovA makes use of our Full Body Interaction framework (FUBI) [29]. In FUBI postures and gestures are defined in an XML-based definition language and are detected by a motion tracking device, such as the Microsoft Kinect. For NovA we defined a set of behavioral primitives, which includes relations between hands, elbows, and feet to each other and to other body parts. Further we investigate the torso and head orientation. Based on the behavioral primitives, we implemented the following recognizers:

1. Typical hand positions: hands together at a particular height of the body, neck touch with left/right hand, head touch with left/right hand, head touch with both hands

2. Characteristic Leg Configurations: standing or sitting with legs apart, closed or crossed

3. Characteristic Arm Configurations: standing or sitting with spread arms at a particular height of the body, arms close to body at a particular height, arms stemmed in hips and arms behind the head with elbows facing outward

4. Common Postures for the Upper Trunk: leaning forward and leaning backward

5. Typical head postures & gestures: looking away, head shakes, head nods, head tilts

4.2 Movement Expressivity

In addition to a mechanism for the detection of postures and gestures, NovA provides measurement for the quality of postures and gestures in terms of expressivity features that are computed frame-by-frame. Based on the work by Wallbott [30] and Caridakis et al. [15], we computed the following expressivity features:

Energy/Power (EN) represents the dynamic properties of a movement (e.g. weak versus strong). It is calculated from the first derivative of the motion vectors in all three dimensions where $\boldsymbol{m}()$ is the motion of the specified joint relative to the torso joint and n is the number of frames that are considered for the calculation.

$$EN = \sqrt{\sum_{i=0}^{n}((\boldsymbol{m}(i).x^2 + \boldsymbol{m}(i).y^2 + \boldsymbol{m}(i).z^2)/3)/n}$$

Fluidity (FL) differentiates smooth movements from jerky ones. This feature aims to capture the continuity between movements. It is calculated as the sum of the variance of both hands' motion vectors' norms $(\boldsymbol{l}, \boldsymbol{r})$ (respectively feet for leg postures).

$$FL = Var(\sum_{i=0}^{n} \boldsymbol{l}(i)/n) + Var(\sum_{i=0}^{n} \boldsymbol{r}(i)/n)$$

Spatial extent (SE) is modeled as the space that is used for gesturing in front of the recorded person. It is calculated as the maximum Euclidean distance of the position of the two hands (l ,r) (respectively feet for leg postures).

$$SE = max(d(\mid r(i) - l(i) \mid))$$

Overall activation (OA) represents the quantity of the movement (passive versus active). It is calculated as the sum of the motion vectors' norm of both hands (respectively feet for leg postures):

$$OA = \sum_{i=0}^{n} \mid \boldsymbol{r}(i) \mid + \mid \boldsymbol{l}(i) \mid$$

Temporal extent (TP) represents the duration of a gesture (short vs sustained). The duration of each gesture is computed from the starting and end points synchronized with the recording time in the SSI framework.

4.3 Facial Expressions

For detecting facial expressions, we make use of Fraunhofer Institute's SHORE [31] as well as the Kinect Facial Tracking SDK[1] which are also integrated in our system. Various occurrences of facial expressions, such as smiles, are computed using a threshold based approach.

5 Mapping Social Cues to User States

To determine the user states described in Section 3 NovA supports the use of Bayesian networks. These networks can be assigned directly to the SSI framework for frame-wise updating with probabilities and evidences in real time. A typical network designed to recognize the user states of *Social Attraction*, *Engagement*, *Self Efficacy* and *Attitude* consists of several unconditional, observed nodes that describe the evidences and probabilities monitored by social cue recognition and expressivity calculations and are updated every frame by the framework. These evidence nodes feed into conditional nodes that estimate a higher level statement based on the recognized cues. Observed as well as conditional nodes lead to the final child node, which models a user state. The Bayesian networks can be modeled with existing tools, such as GeNIe[2].

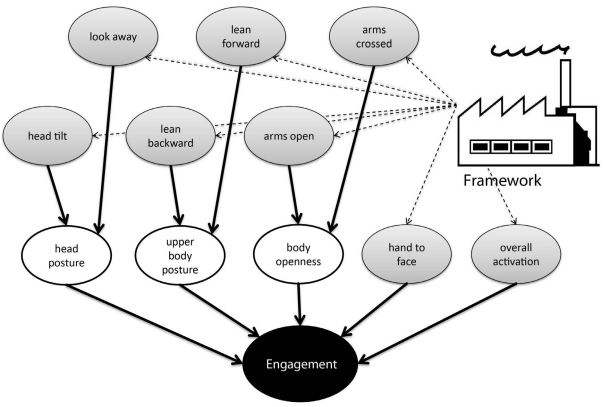

Fig. 2. A simplified Bayesian network to determine Engagement

[1] http://msdn.microsoft.com/en-us/library/jj130970.aspx
[2] http://genie.sis.pitt.edu/

As an example, Figure 2 shows the diagram of a simplified Bayesian network, meant to recognize *Engagement*. The structure of the network has been defined in accordance to social theories, extracted from relevant literature (See Section 3). Eight observed, unconditional parent nodes (head tilt, look away, lean backward, lean forward, arms open, arms crossed, hand to face and overall activation) are constantly updated by the framework with evidences from gesture recognizers and calculation values from expressivity calculation. Six of these parent nodes influence interconnected conditional nodes (head posture, upper body posture and body openness), the other two directly feed into the final *Engagement* node. We want to point out that in Figure 2 we demonstrate a simplified version of a predefined Bayesian network to determine *Engagement*. Besides predefined BNs, starting probabilities as well as probability tables and related social cues can also be defined by users of the system themselves.

6 The User Interface of NovA

6.1 Graphical Interface

The user interface of NovA has been developed following the requirements of tools for annotating human social interactions. Like other tools, it offers annotation on multiple tracks based on a user-defined coding scheme that has been, in our case, adapted to the situation of human-human or human-agent dialogue.

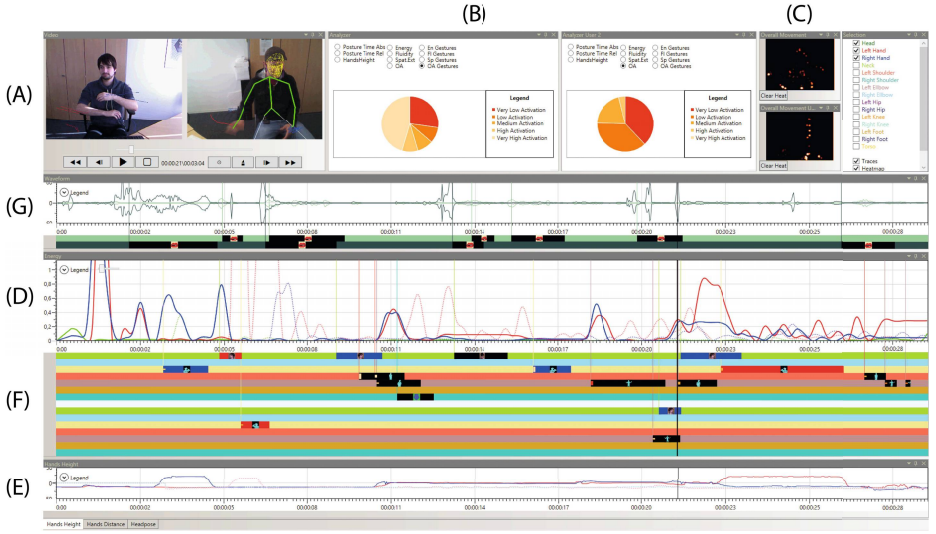

Fig. 3. NovA's graphical user interface. In this instance data for two users has been loaded. It shows both videos (with and without skeleton, (Figure 3 A)), pie charts for expressivity features (Figure 3 B), heatmaps (Figure 3 C), a waveform graph with voice activity detection events (Figure 3 G), the timeline graph showing automatically created annotations (Figure 3 F) and the hands height graph (Figure 3 E).

Annotations are automatically added corresponding to social cues recognized by the system by placing boxes along a horizontal timeline. Typically, different kinds of behaviors are coded on different parallel tracks so that their temporal relationships are clearly visible.

Figure 3 shows the graphical user interface of NovA. The GUI includes the video recordings of up to two recorded people (Figure 3 A) in addition to diagrams with descriptive statistics (Figure 3 B), Heatmaps (Figure 3 C) as well as timeline diagrams showing the temporal dynamics of their behaviors.

Timeline diagrams contain:
1. Tracks that correspond to behavioral characteristics collected frame by frame, such as motion energy (Figure 3 D) or the height of the hands (Figure 3 E) and
2. Tracks that correspond to events, such as the occurrence of particular social cues (Figure 3 F).

The screenshot shown in Figure 3 only depicts a particular instantiation of the graphical user interface of NovA which can be dynamically adapted depending on the tasks to be conducted. For example, all windows are customizable in size and position and can also be removed or hidden.

Another example of a frame-by-frame annotation is the representation of the waveform corresponding to the audio signal (Figure 3 G). Phases with high peaks indicate high intensity (e.g. a loud voice) while phases with tiny peaks represent silent phases.

Each annotation includes additional information that may be displayed on demand. It includes a reference picture, the exact duration, calculated expressivity parameters and a description of a possible interpretation. For the case of detection errors or the need of adding annotations that could not be detected automatically, NovA also offers the possibility to manually add or delete event-based annotations or to edit their temporal position and duration. In addition, annotation schemes are fully customizable and can contain both automated and manual annotations.

6.2 Illustrating Example

In the following, we present an example to illustrate behavior analysis in NovA. Figure 4 illustrates how the system tries to determine the level of engagement of an interviewee. On the left picture (a) the participant has an open body posture, while looking towards the interlocutor and orientating his body in the same direction. In the center picture (b), nothing specific is detected, and the right picture (c) demonstrates the outcome when the participant uses body language regarded as indicator of a low amount of engagement (see Section 3), such as leaning back, looking away and crossing the arms. Bar charts are representing the outcome of the user state recognition for each calculation, which is performed every second.

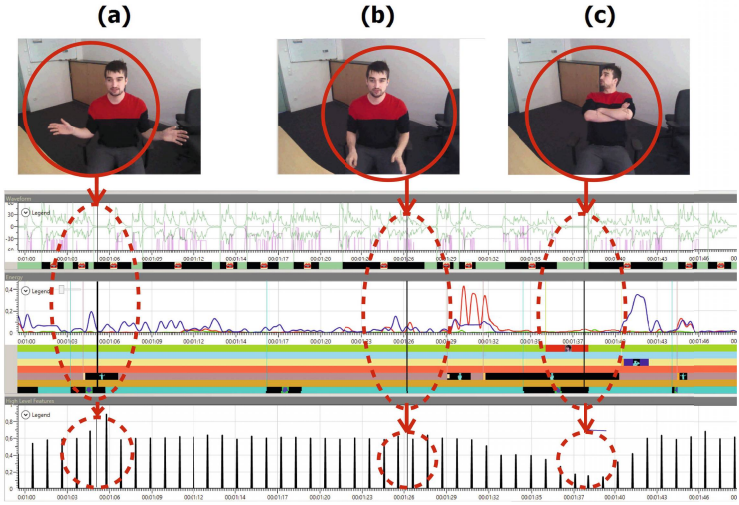

Fig. 4. Comparison of detected cues for high (a), medium (b) and low engagement (c)

7 Applications

NovA was originally developed for the EU-funded project TARDIS[3]. TARDIS attempts to support young adults in job interviews by developing a scenario-based game with virtual agents acting as recruiters. One large issue Europe faces is the rising number of young people who are out of employment, education or training (NEETs). NEETs often have underdeveloped socio-emotional and interaction skills [32], such as a lack of self-confidence, lack of sense of their own strengths or social anxiety [33]. This can cause problems in various critical situations such as job interviews where they need to convince the recruiter of their fit in a company. To address this issue, many European countries have specialized inclusion centers meant to aid young people secure employment through coaching by professional practitioners. One problem of this approach is that it is very expensive and time-consuming. Considering this, technology-enhanced solutions, such as digital games, present themselves as viable and advantageous alternatives to the existing human-to-human coaching practices.

Within TARDIS, NovA serves to analyze the learner's social cues when interacting with a virtual recruiter during a job interview simulation. Providing such an environment is highly desirable from the point of view of improving practice, since it enables a repeatable experience that can be modulated to suit the individual needs of the learner. It may also mitigate negative side effects resulting from real-life settings, in particular, the stress associated with engaging in unfamiliar interactions with others. In this context, the recognition rate of our social cue detection for body language and voice activity detection has been

[3] Training young Adult's Regulation of emotions and Development of social Interaction Skills - `http://tardis.lip6.fr`

Fig. 5. NovA serves to analyze the learner's social cues when interacting with a virtual recruiter during a job interview simulation in the TARDIS project

evaluated and achieved a mean recognition rate of 88% [34]. Various knowledge elicitation studies have been conducted using real job seeking youngsters and trainer practitioners [35]. Ongoing user experience evaluations have shown so far that user's self reports about their behavior characteristics correspond to NovA's calculated results. The data of mentioned studies has been used to shape the design of NovA, the choice in social cues and our preliminary Bayesian networks (BNs) for detecting user states.

8 Conclusion

In this paper we presented NovA, a system that processes data from state of the art sensor technology for automated recognition and analysis of human behavior. NovA offers an online component for recording and processing data in real time and a user interface to visualize and analyze data. The user interface's primary use is the debriefing and post-hoc analysis of social interactions. It allows users to reflect on their behavior, and thus learn to perform better in social situations.

NovA is able to automatically recognize various social cues, such as gestures, postures, expressivity features or facial expressions. Additionally NovA provides support for feeding such social cues to specially designed Bayesian Networks to compute higher level user states.

The possibility to automatically detect human behavior can help researchers from various disciplinary areas. On the one hand engineers might use higher-level information on the user's behavior to improve assistant robots, virtual agents or other user assistant software such as the job-interview training game developed in the TARDIS project. On the other hand automated behavioral analysis- and coding can help psychological researchers by reducing their workload. To maximize our contribution to the research community we made NovA available for download at http://openssi.net/nova.

As part of our future work we plan to further validate the performance of the system's user state detection. Additionally, we plan to investigate the use of physiological sensor devices and eye-trackers that promise more detailed information on a user's affective and emotional state.

Acknowledgments. This work was partially funded by the European Commission within FP7-ICT-2011-7 (Project TARDIS, grant agreement no. 288578).

References

1. Mehrabian, A.: Silent messages: Implicit Communication of Emotions and Attitudes. Wadsworth Publishing Co Inc., Belmont (1981)
2. Eagle, N., Pentland, A.: Reality mining: Sensing complex social signales. J. of Personal and Ubiquitous Computing 10(4), 255–268 (2006)
3. Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H.: Elan: A professional framework for multimodality research. In: Proc. of the Fifth International Conference on Language Resources and Evaluation (LREC), pp. 879–896 (2006)
4. Kipp, M.: Anvil: The video annotation research tool. In: Handbook of Corpus Phonology. Oxford University Press, Oxford (2013)
5. Schmidt, T.: Transcribing and annotating spoken language with exmaralda. In: Proc. of the LREC-Workshop on XML Based Richly Annotated Corpora, Lisbon 2004, pp. 879–896. ELRA, Paris (2004)
6. Curhan, J., Pentland, A.: Thin slices of negotiation: predicting outcomes from conversational dynamics withing the first 5 minutes. J. of Applied Psychology 92(3), 802–811 (2007)
7. Pentland, A.: Automatic mapping and modelling of human networks. Physica A(378), 59–67 (2007)
8. Schuller, B., Müller, R., Eyben, F., Gast, J., Hörnler, B., Wöllmer, M., Rigoll, G., Höthker, A., Konosu, H.: Being bored? recognising natural interest by extensive audiovisual integration for real-life application. Image Vision Comput. 27(12), 1760–1774 (2009)
9. Rich, C., Ponsleur, B., Holroyd, A., Sidner, C.L.: Recognizing engagement in human-robot interaction. In: Proc. of the 5th ACM/IEEE Intl. Conf. on Human-Robot Interaction, HRI 2010, pp. 375–382. IEEE Press, Piscataway (2010)
10. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. 31(1), 39–58 (2009)
11. Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., Zancanaro, M.: Multimodal recognition of personality traits in social interactions. In: Proc. of the 10th International Conference on Multimodal Interfaces, ICMI 2008, pp. 53–60. ACM, NY (2008)
12. Dong, W., Lepri, B., Cappelletti, A., Pentland, A.S., Pianesi, F., Zancanaro, M.: Using the influence model to recognize functional roles in meetings. In: Proc. of the 9th International Conference on Multimodal Interfaces, ICMI 2007, pp. 271–278. ACM, New York (2007)
13. Hung, H., Gatica-Perez, D.: Estimating cohesion in small groups using audio-visual nonverbal behavior. Trans. Multi. 12(6), 563–575 (2010)
14. Sandbach, G., Zafeiriou, S., Pantic, M., Yin, L.: Static and dynamic 3d facial expression recognition: A comprehensive survey. Image Vision Comput. 30(10), 683–697 (2012)
15. Caridakis, G., Raouzaiou, A., Karpouzis, K., Kollias, S.: Synthesizing gesture expressivity based on real sequences. In: Workshop on Multimodal Corpora: from Multimodal Behaviour Theories to Usable Models. LREC, Genoa (2006)
16. Vogt, T., André, E., Bee, N.: Emovoice - a framework for online recognition of emotions from voice. In: André, E., Dybkjær, L., Minker, W., Neumann, H., Pieraccini, R., Weber, M. (eds.) PIT 2008. LNCS (LNAI), vol. 5078, pp. 188–199. Springer, Heidelberg (2008)

17. Kleinsmith, A., Bianchi-Berthouze, N.: Form as a cue in the automatic recognition of non-acted affective body expressions. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011, Part I. LNCS, vol. 6974, pp. 155–164. Springer, Heidelberg (2011)
18. Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. IEEE Trans. Pattern Anal. Mach. Intell. 30(12), 2067–2083 (2008)
19. Camurri, A., Volpe, G., De Poli, G., Leman, M.: Communicating expressiveness and affect in multimodal interactive systems. IEEE MultiMedia 12(1) (2005)
20. Scherer, S., Marsella, S., Stratou, G., Xu, Y., Morbini, F., Egan, A., Rizzo, A(S.), Morency, L.-P.: Perception markup language: Towards a standardized representation of perceived nonverbal behaviors. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS, vol. 7502, pp. 455–463. Springer, Heidelberg (2012)
21. Simpson, J.A., Harris, B.A.: Interpersonal attraction. In: Weber, A.L., Harvey, J.H. (eds.) Perspectives on Close Relationships, pp. 45–66. Prentice Hall (1994)
22. McGinley, H., LeFevre, R., McGinley, P.: The influence of a communicator's body position on opinion. J. of Personality and Social Psychology 31(4), 686–690 (1975)
23. Schouwstra, S., Hoogstraten, J.: Head position and spinal position as determinants of perceived emotional state. Perceptual and Motor Skills 81, 673–674 (1995)
24. Sidner, C.L., Kidd, C.D., Lee, C., Lesh, N.: Where to look: a study of human-robot engagement. In: IUI 2004: Proc. of the 9th International Conference on Intelligent User Interfaces, pp. 78–84. ACM Press, New York (2004)
25. Pease, A.: Body Language. Sheldon Press, London (1988)
26. Bandura, A.: Self Efficacy: The Exercise of Control. Palgrave Macmillan, New York (1997)
27. Forgas, J.P., Cooper, J., Crano, W.D.: The Psychology of Attitudes and Attitude Change. Taylor & Francis Group, New York (2010)
28. 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, Barcelona, Spain (2013)
29. Kistler, F., Endrass, B., Damian, I., Dang, C., André, E.: Natural interaction with culturally adaptive virtual characters. Germany Journal on Multimodal User Interfaces Heidelberg/Berlin (2012)
30. Wallbott, H.: Bodily expression of emotion. European Jrl. of Social Psychology (28), 879–896 (1998)
31. Ruf, T., Ernst, A., Küblbeck, C.: Face detection with the sophisticated high-speed object recognition engine (shore). In: Microelectronic Systems, pp. 243–252. Springer (2011)
32. Hammer, T.: Mental health and social exclusion among unemployed youth in scandinavia. a comparative study. Intl. Jrnl. of Social Welfare 9(1), 53–63 (2000)
33. Pan, X., Gillies, M., Barker, C., Clark, D.M., Slater, M.: Socially anxious and confident men interact with a forward virtual woman: An experiment study. PLoS ONE **7**(4) (2012) e32931
34. Damian, I., Baur, T., André, E.: Investigating social cue-based interaction in digital learning games. In: Proc. of the 8th International Conference on the Foundations of Digital Games, SASDG (2013)
35. Porayska-Pomsta, K., Anderson, K., Damian, I., Baur, T., André, E., Bernardini, S., Rizzo, P.: Modelling users' affect in job interviews: Technological demo. In: Carberry, S., Weibelzahl, S., Micarelli, A., Semeraro, G. (eds.) UMAP 2013. LNCS, vol. 7899, pp. 353–355. Springer, Heidelberg (2013)