Modeling grounding for interactive social companions

Gregor Mehlmann, Kathrin Janowski, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Mehlmann, Gregor, Kathrin Janowski, and Elisabeth André. 2016. "Modeling grounding for interactive social companions." *KI - Künstliche Intelligenz* 30 (1): 45–52. https://doi.org/10.1007/s13218-015-0397-5.

Modeling Grounding for Interactive Social Companions

Gregor Mehlmann · Kathrin Janowski · Elisabeth André

Abstract Grounding is an important process that underlies all human interaction. Hence, it is also crucial for social companions to interact naturally. Maintaining the common ground requires domain knowledge but has also numerous social aspects, such as attention, engagement and empathy. Integrating these aspects and their interplay with the dialog management in a computational interaction model is a complex task. We present a modeling approach overcoming this challenge and illustrate it based on some social companion applications.

1 Introduction

We recently observe a growing interest in the development of artificial social companions, such as sociable and collaborative robots or embodied conversational agents. They serve as playmates [7], trainers and coaches [29, 12] or assistants in health and elderly care [16]. They are expected to offer a natural and intuitive social interaction and to be adaptive to the users' interests and needs. An essential prerequisite to reach these goals is to maintain the *common ground*, "the set of knowledge, beliefs and suppositions that the participants believe they share" [10], during an interaction. We use the term *grounding* for the sum of all behavioral aspects involved in establishing, maintaining and repairing the common ground with the *least collaborative effort* [11].

Gregor Ulrich Mehlmann
Human Centered Multimedia
Augsburg University, Germany
E-mail: mehlmann@hcm-lab.de

Grounding is a reciprocal process, that means it does not only include the production of multimodal behaviors to achieve common ground, but also monitoring the interlocutor's behaviors and appropriate interventions if the common ground is disrupted. Simulating multimodal grounding in a social companion is a challenge due to the complexity and heterogeneity of the processes that need to be tightly coordinated. Small discrepancies in the synchronization of grounding behaviors may make the companion appear unnatural and awkward. Since the individual behavioral aspects are highly interwoven, they cannot be treated in isolation, but much rather call for a uniform modeling approach.

In this paper, we present a computational interaction framework for social companions that handles multimodal grounding aspects within a uniform model and integrates them with dialog management. Our framework is based on an expressive and adaptive modeling formalism that keeps the model manageable, extensible and reusable. In the following, we first review a variety of multimodal grounding processes that need to be implemented by the model including the disambiguation of dialog utterances, the fluent exchange of information and participant roles, the display of cognitive and emotional states as well as processes for establishing and maintaining engagement. After an identification of requirements and challenges and a discussion of related work, we present our modeling approach for handling all these grounding phenomena within a single interaction framework and illustrate it with examples based on our recently developed social companion applications.

2 Grounding

We now review the aspects of grounding that need to be mastered by social companions to interact naturally.

Disambiguation and Clarification Dialogs
Face to face interaction does not take place in an empty space, but is linked to the situative environment. When referring to objects in this environment, humans distribute information across different modalities, depending on the effort and the expressive power of each channel, and rely on their partners' ability to combine this information in order to resolve ambiguities [27]. Referential grounding requires to establish and maintain *joint attention* using gaze and gestures to monitor and direct the attention of others to objects, actions and events [24]. For example, the speaker's gaze behaviors that are aligned with spoken utterances may help the listener discriminate the meant objects from alternatives. Vice versa, the speaker may detect a disruption of the common ground by monitoring the listener's gaze behavior and apply appropriate repair mechanisms. In case the common ground has been lost, a clarification dialog is initiated in which disambiguities are resolved.

Exchanging Information and Participant Roles
Reaching common ground within a conversation also requires a mechanism for regulating the allocation of speaker and listener roles [18,14,10]. Thereby, the precise alignment of speech and gaze serves as a key signal in managing or inhibiting the exchanges of these roles. Speakers usually look away from their addressees to indicate that they want to keep the floor and look at one of their partners at the end of their utterance to pass the floor to this participant [25]. The exchange of turns can be delayed if a contribution does not end with such a mutual gaze at a partner [18]. Listeners use *backchannels* to let the speaker know that they understand what has been said [30,2], thus grounding the information states without requesting the turn. In return, speakers occasionally perform a short glance of mutual gaze to the addressee without yielding the turn with the aim to elicit backchannels at specific points in time [18,2,5].

Revealing Cognitive and Emotional States
Grounding requires being aware of the partners' cognitive and emotional states while producing adequate verbal and nonverbal cues to reveal one's own thoughts and feelings [18,3,13]. Thereby, gaze cues and facial expressions are among the key signals to display cognitive operations and emotional states. For example, speakers usually avert gaze when planning speech [18] or updating beliefs, desires, and intentions [13]. Both listeners and speakers try to establish mutual gaze when they signal and ensure understanding by continually producing and eliciting acoustic and visual backchannels [18, 30,2,5]. They convey emotional states via a variety of modalities, such as facial displays and postures that are frequently enhanced by gaze behaviors. For example, avoidance-oriented emotions, such as fear, are typically accompanied by averted gaze while directed gaze is linked to approach-oriented emotions, such as joy [1].

Engagement and Emotional Contagion
Grounding is also closely related to *engagement*, a process "by which two or more participants establish, maintain and end their perceived connection during interactions they jointly undertake"[28]. Firstly, a companion should show engagement by producing appropriate backchannels and attentive behaviors. Vice versa, it has to monitor the interlocutor's level of engagement by analyzing the latter's behavioral cues and intervene if necessary. Sharing and understanding each other's emotions may further increase engagement in an interaction. A simple form of emotional grounding can be realized by mimicking the emotional cues of the conversational partner. Such ideomotoric behaviors may convey the impression of empathy even in the absence of any understanding of the other's emotional state. Usually, empathy requires, however, a deeper level of mind reading [4] from the companion. This means the companion has to appraise the situation from the perspective of the conversational partner to produce a sensitive response.

3 Challenges

We now identify requirements and challenges for a modeling approach to cope with the aspects of grounding.

Incremental and Multimodal Processing
Most aspects of grounding include the multimodal fusion of events and the incremental recognition of the user's multimodal behavior patterns. This requires the evaluation of various temporal and semantic relations between events that may carry information ranging from unprocessed data, such as eye gaze coordinates, to more

abstract forms, such as fully interpreted dialog acts. In symmetry, many aspects also enfold the stepwise generation and multimodal fission of the agent's behavior. A formalism for multimodal fission must allow to specify the same relations for the alignment of speech and nonverbal behaviors when generating the agent's behavior.

Concurrent and Interleaving Processes
Grounding involves various processes for perception, interpretation, decision-making and behavior control. For example, the computation of the user's gaze shifts to specific objects based on eye tracking data and domain knowledge is a dedicated process. Other processes are combining events to recognize complex behavioral patterns on higher processing stages. For example, the disambiguation of verbal references with aligned gaze fixations represents such a process. Finally, the different reactive and deliberate multimodal behavioral aspects can be understood as individual parallel processes. These numerous concurrent processes are tightly interleaved and have to be properly coordinated and synchronized.

Priorities of Behavior and Resumption
At certain points during an interaction some aspects of grounding become more important than others. For example, a listener usually divides his gaze between the speaker and the other participants. However, it is more important for grounding that he follows the speaker's occasional gaze movements to objects. A sudden distraction, in turn, must interrupt the gaze following behavior which, however, should be resumed afterwards. The modeling formalism must allow to express priorities between processes that execute conflicting aspects of grounding. Each process must remember the point of its interruption to coherently resume its execution.

4 Related Work

Most related work focuses on computational models for individual aspects of grounding, such as engagement [17], cognitive states [21] or turn-taking [8] in isolation. These models do not cope with the tight alignment or the complex interplay and coordination of the numerous aspects of grounding that we presented in Section 2. More fundamental related research aims at the development of generic and versatile modeling languages for multimodal fusion [20], dialog management [26] and interaction management [15]. They tackle individual challenges of those identified in Section 3, without focusing on grounding or similar concepts, at all. Our work goes beyond this state-of-the-art because we present a uniform modeling approach that handles all those challenges [22] and use it to develop a computational interaction model that enfolds all the aspects of grounding that have so far been regarded only in isolation [23].

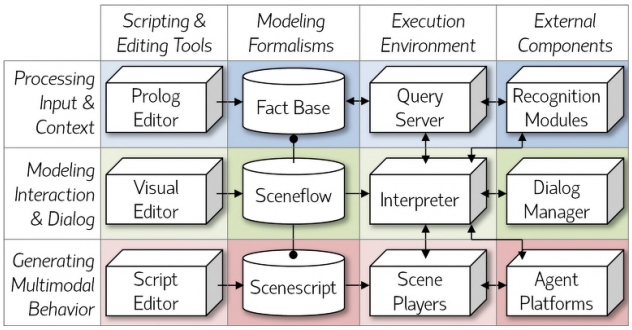

Fig. 1 VSM's software architecture and modeling processes.

5 Modeling

Our social companion applications have been developed with the *VisualSceneMaker* authoring tool [15,22]. As shown in Figure 1, *VSM* divides the authoring task into three subtasks relying on visual and declarative modeling languages. This facilitates iterative prototyping, distributed development and the creation of clearly structured, easily maintainable and reusable interaction models. It has an IDE[1] that enables the visual editing of a model and the real-time visualization of its execution.

Processing User Input and Context
User input events and domain knowledge are uniformly represented as *feature structures* [9] and maintained in a *fact base* that is implemented in *Prolog*. Input events are usually preprocessed by modality-specific interpretation modules before they are asserted to the fact base. They carry timestamps and confidence values as well as modality-specific semantic information, such as gaze targets, recognized gestures or parsed dialog acts.

The fact base contains predefined and user-defined logic predicates for reasoning on domain knowledge and inferring data from user input events. They are used to compute structural and functional constraints as well as temporal, spatial, ordering and semantic relations for multimodal fusion and behavior pattern recognition.

[1] http://scenemaker.dfki.de/

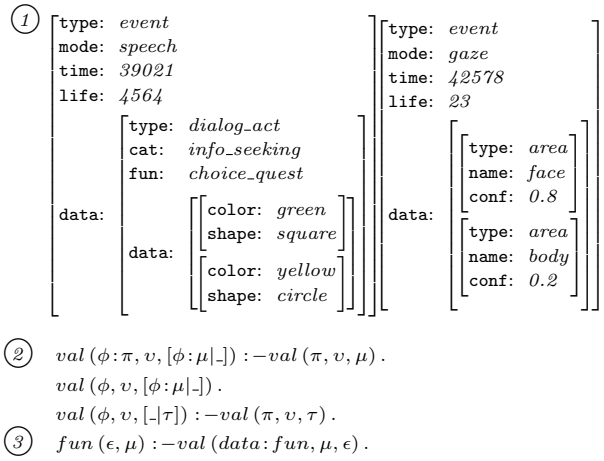

Fig. 2 Exemplary feature structures and logic predicates.

For example, Figure 2 shows two feature structures (Fig. 2 ①) representing, first, a dialog act parsed from the user's choice question *"the green square or the yellow circle?"* in our *RobotPuzzle* application (Fig. 6 Ⓐ) and, second, a gaze event containing probabilities with which the user looks at certain body parts of the social companion in our *SocialCoach* application (Fig. 6 Ⓑ). Below is the definition of the predefined predicate $val/3$ (Fig. 2 ②) used to extract a feature value from such a structure and the predicate $fun/2$ (Fig. 2 ③) used to infer feature *data:fun* of a dialog act based on $val/3$.

Specifying Behavior and Dialog Content

Multimodal behavior and dialog content are specified in a set of *scenes* organized in a *scenescript*. A scenescript contains dialog utterances and commands for nonverbal behavior such as gestures, postures, gaze and facial expressions. It may provide a number of variations for each scene in order to avoid repetitive behavior. It may be manually scripted for the purpose of rapid prototyping or automatically generated from context knowledge.

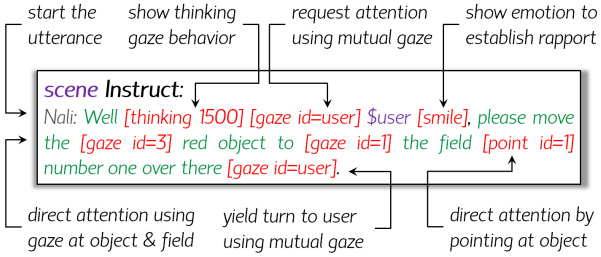

Fig. 3 A scene with spoken text and nonverbal behavior.

Figure 3 shows an exemplary scene from the *RobotPuzzle* application (Fig. 6 Ⓐ) in which several aspects of grounding are realized. The scene contains a spoken utterance aligned with nonverbal behaviors for revealing cognitive and emotional states and gaze cues for directing the user's attention and regulating the floor. While *VSM* supports the usage of more expressive behavior modeling languages, such as *BML* [19], our experience has shown that this format is usually sufficient for the alignment of speech and nonverbal behaviors.

Modeling Interaction and Dialog Flow

Interaction logic and dialog flow are modeled with a *sceneflow*, a domain-specific hierarchical and concurrent state chart. It is used to control and synchronize parallel processes modeling the agents' behavior and input processing on different abstraction levels. A sceneflow consists of different types of *nodes* and *edges* that may be annotated with statements of a scripting language. A *basic node* may contain type- and variable definitions, assignments and calls to functions of the underlying plug-in modules shown in Figure 1, such as scene playback commands or queries to the logic fact base. Local variables are used unbound or instantiated within queries in order to exchange data between the sceneflow and the fact base [22]. A *super node* contains nested parallel sceneflows whose execution start points are defined via *start nodes*. Parallel sceneflows can be synchronized via shared variables and exchange events via the fact base. Branching strategies are realized with different types of edges that may be taken probabilistically, after some timeout or under some condition in the form of a conditional expression or a logic query.

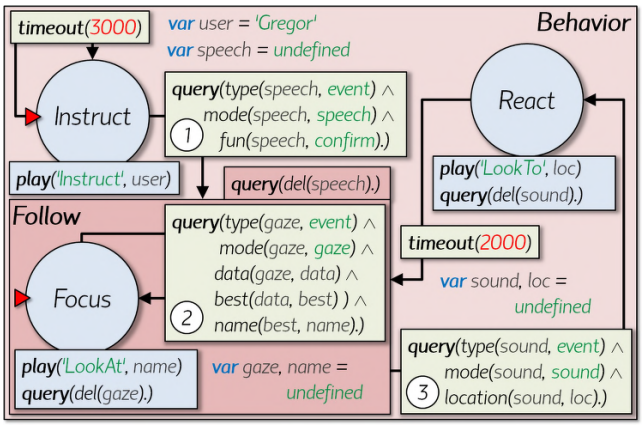

Fig. 4 A sceneflow modeling some simplified gaze behavior.

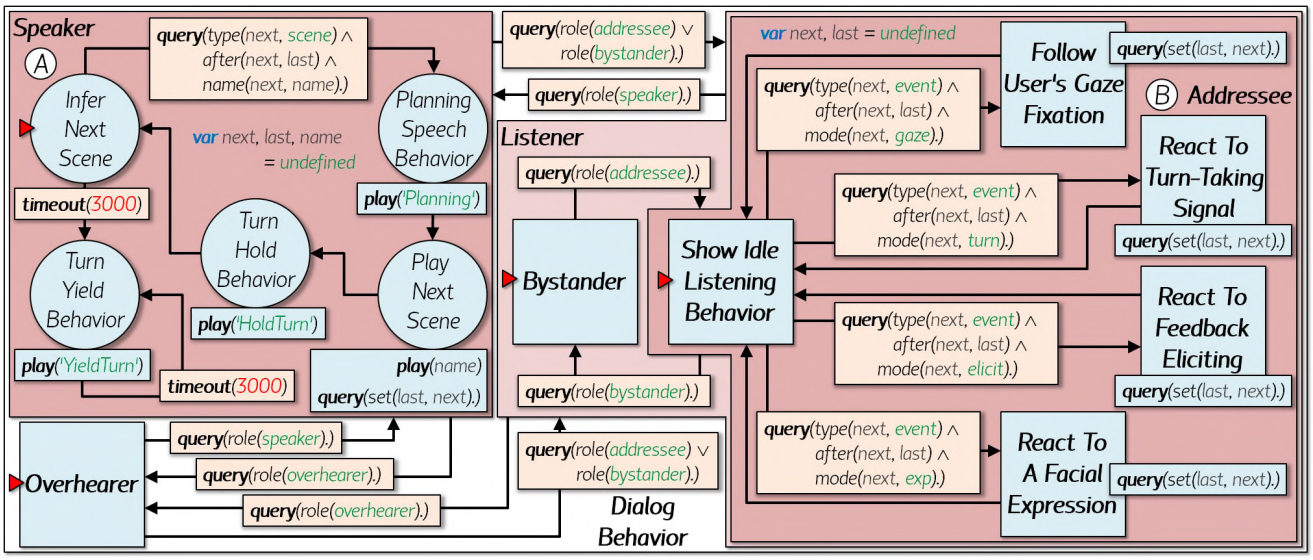

Fig. 5 A simplified version of the process modeling the dialog behavior of social companions in the behavior control layer.

Figure 4 shows a simplified part of the sceneflow used in the *RobotPuzzle* application (Fig. 6 Ⓐ). It consists of the super node *Behavior* whose execution starts at start node *Instruct*, marked with the red triangle. Here we play the scene *Instruct* from Figure 3 before taking the timeout transition after 3 seconds. If the user confirms the instruction before, for example by saying *"Okay"*, we detect a dialog act with function *confirm* in the fact base. In this case we take transition ① to super node *Follow* and delete the according speech event. The query guarding transition ① uses, among others, the predicate $fun/2$ from Figure 2. Super node *Follow* starts with start node *Focus* in which we successively await the user's new gaze fixation events in the fact base. Whenever we detect a new gaze event, we take transition ② whereby we extract the name of the focused object. Then we follow the user's gaze to this object by playing the scene *LookAt* with the object's name as argument and delete the consumed gaze event. Whenever a disrupting sound event is detected, we immediately extract its location while taking transition ③ to node *React*. Thus, we stop following the user's gaze and look at this location in scene *LookTo* before we resume the gaze following behavior after 2 seconds.

This illustrative example shows that the stepwise execution of sceneflows allows for the close interleaving of input processing and behavior generation. The hierarchical refinement of sceneflows may be used to realize priorities between several conflicting behavior aspects.

6 Realization

We used the same interaction model architecture for all our social companion applications. It divides the model into different layers of parallel and nested sceneflows that are coordinated via events. They model the processes for perception, interpretation, decision-making and behavior control that we mentioned in Section 3.

Different Processing Layers
On the *perception* level, we preprocess modality-specific input and context changes, for example when computing the user's gaze shifts. On the *integration* level, we combine events of different modalities based on temporal and semantic constraints, for example when recognizing the coordinated use of speech and gaze for turn-taking signals or feedback eliciting cues. On the *decision* layer, all information comes together and is used to make decisions that control the behavior. For example, we decide if an agent may be interrupted and assign the participant roles. The *behavior* layer contains processes executing the various multimodal behavioral aspects.

Regulating Participant Roles
The *behavior* layer enfolds the process that is modeling the *dialog behavior* as simplified depicted in Figure 5. It is switching between the participant roles based on the role decisions made in the layers below. They depend on, firstly, the recent turn-regulation signals of all interaction partners, secondly, a joint policy determining

which participant may interrupt another, and, finally, whether the agent wants to contribute to the dialog.

The agent's dialog behavior starts in the *overhearer* role in which it is not participating in an interaction and is showing idle behavior. In the *bystander* role it occasionally observes and listens to an interaction between two or more other participants but does not make a dialog contribution. When it becomes the *addressee* then it is actively listening to the speaker and may request speaking turns or contribute to the conversation.

Producing Grounding Behaviors

Most aspects of grounding are modeled via the interplay of multiple processes in the different layers of the model. The control and the production of the resulting expressive behavior is up to the dialog behavior process.

While the agent is in the speaker role (Fig. 5 Ⓐ), most grounding behaviors, such as revealing the cognitive state, directing the user's attention or establishing mutual gaze for engagement, are realized in scenes. As shown in Figure 3, the scene then contains the appropriate specifications for the agent's nonverbal behavior. Universal grounding behaviors, such as gaze aversion while planning and gaze to the addressee when starting an utterance as well as turn-regulation signals following an utterance, may be hard-wired within the sceneflow.

During the addressee (Fig. 5 Ⓑ) and the other listener roles the agent produces a role-dependent default behavior and shows grounding behaviors by reacting to emotional expressions and turn-taking, feedback eliciting or attention direction signals of the other agents before resuming the default behavior. This happens in dedicated nested sceneflows that may be adapted to implement specific strategies or variations of these grounding behaviors for a particular companion or application.

7 Applications

Figure 6 shows social companion applications that we developed to validate our modeling approach presented in Sections 5 and 6. Although these companions perform different tasks and have varying technical capabilities, their interaction behavior is successfully controlled with our framework. We used these applications to research different aspects of grounding in some experiments. Our framework's ability to log context parameters during the execution of an interaction model considerably eased the evaluation of these experiments.

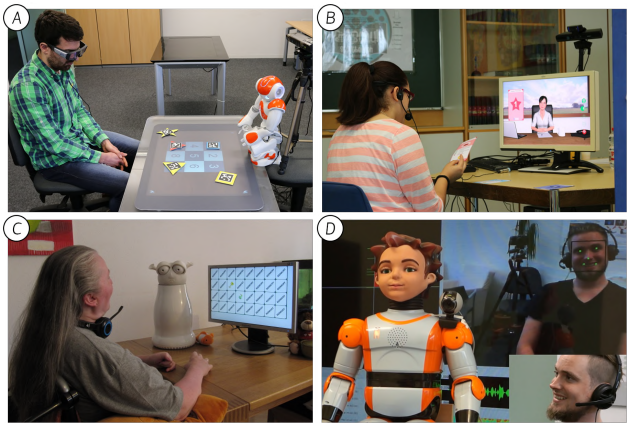

Fig. 6 Social companion applications developed at our lab.

A Collaborative Social Robot Instructor

Figure 6 Ⓐ shows our *RobotPuzzle* application that we use to study the roles of gaze for grounding. It is a collaborative game between the user, wearing an eye-tracker and a headset, and a *NAO*[2] robot on a *Microsoft*[3] Surface table. The robot guides the user in sorting puzzle pieces into the correct slots on the table. The puzzle pieces have distinguishable features such as a shape, size, color and position. User and robot may use a combination of gaze and speech to refer to specific objects, to yield and take the turn and to draw the other's attention to objects or themselves. Thereby, they may produce ambiguous references that can then be resolved by multimodal disambiguation, combining gaze and speech, or a clarification dialog. An evaluation showed that gaze following to referred objects and the multimodal disambiguation of references using gaze make the interaction more pleasant and efficient [23].

A Coach for Simulated Job Interviews

Figure 6 Ⓑ shows our *JobCoach* application with a virtual *Charamel*[4] character in the role of an interactive virtual recruiter. Young unemployed or uneducated people can practice their socio-emotional skills and stress management in simulated job interviews. The character monitors their verbal and non-verbal behavior during the interaction and gives advice to them on how to improve their performance, such as showing a friendly smile when introducing themselves. A user study with pupils demonstrated clear benefits of the

[2] http://www.aldebaran-robotics.com/
[3] http://www.microsoft.com/
[4] http://www.charamel.com/

experienced-based learning approach with our application over traditional learning methods and showed, that the virtual character helped the pupils to better control negative emotional states, such as nervousness [12].

A Game-Playing Social Robot Companion
Figure 6 © shows our *RoboGames* application with a social game-playing robot similar to the agent described in [7]. A *Robopec Reeti*[5] robot plays different card or board games with the user, for example to train the mental fitness of elderly people living on their own. Our focus here is on the conversational engagement mechanisms and the display of the robot's cognitive and emotional states. These mechanisms include, for example, mutual gaze with the user, comments about unusual delays, and "thinking" behavior such as examining the game screen for options or looking up while remembering the location of a matching card. Furthermore, the robot responds emotionally to various events in the game, both by facial expressions and appropriate comments. An ongoing student project will replace the manually authored emotions with a more sophisticated affect model [6] to enable autonomous reactions based on the robot's given personality. We are thus expecting to make the robot's behavior more credible and to sustain the user's interest over a long period of time.

An Emotionally Sensitive Social Robot
Figure 6 Ⓓ shows our test bed for empathy modeling. Here, the user's mood is inferred from various cues such as their tone of voice and facial expression. This in turn controls the behavior of a *RoboKind R-50 Zeno*[6] which exhibits two distinct empathy mechanisms [6]. First, the robot constantly adapts its facial expression to match the user's emotion, signaling a basic awareness of their situation. Second, the robot verbally expresses happiness or pity for the user. Unlike the direct and seemingly instinctive mirroring, this reflects an active interest in the user's well-being, a key requirement for a social companion. We will combine this mechanism with personalized recommender systems to allow for higher reasoning and constructive advice. For example, the robot might suggest a meeting with friends when the user complains about loneliness or offer to call a doctor when the user is feeling sick. We expect the empathy display to provide additional comfort and encourage the user to take their companion's advice.

[5] http://www.reeti.fr/
[6] http://www.robokindrobots.com

8 Conclusion

We presented a conceptual and technical framework for modeling the grounding behavior of interactive social companions. First, we showed that natural grounding behavior requires the precise synchronization of numerous parallel and bidirectional behavioral aspects. Then, we identified the resulting challenges for an expressive and adaptive behavior modeling formalism. Going beyond previous work, which covers only specific aspects of grounding in isolation, we presented a novel uniform modeling approach that copes with those challenges.

Our approach unifies advantages of statecharts, such as their flexibility and reusability, and logic programming, such as its expressiveness and declarative nature. It relies on a hierarchical and concurrent statechart dialect for interaction control that is enriched with queries to a logic fact base for multimodal fusion and knowledge reasoning. Multimodal dialog content is specified in a simple script format which can be manually created or generated by a dialog manager. The separation of the modeling task into these parallel processes facilitates iterative prototyping and distributed development.

Our approach does not restrict the usage or combination of modalities and allows to express a variety of temporal and semantic constraints. The stepwise execution of the interaction model enables the precise alignment and incremental interleaving of input processing and behavior generation. The parallel decomposition allows to model and synchronize parallel processes for perception, integration, decision-making and behavior control. The hierarchical refinement can be used to realize priorities among behavioral aspects of grounding.

We illustrated and validated our modeling approach by developing various social companion applications that all used the clearly structured, adaptable and reusable interaction model, presented in this paper. We adjusted details of this model to influence various grounding behaviors, such as attention following, turn-taking strategies and emotional competence for the different agents. Our framework's ability to log context data during the execution eased the evaluation of our applications.

Acknowledgements This work has been partially funded by the European Commission within the 7th Framework Programme in the research project *TARDIS*, the European Union's Horizon 2020 Research and Innovation Programme in the research project *KRISTINA* and the German Federal Ministry of Education and Research in the research project *EmpaT*.

References

1. Adams, R.B., Kleck, R.E.: Effects of Direct and Averted Gaze on the Perception of Facially Communicated Emotion. Emotion **5**(1), 3–11 (2005)
2. Allwood, J., Nivre, J., Ahlsén, E.: On the Semantics and Pragmatics of Linguistic Feedback. Journal of Semantics **9**(1), 1–26 (1992)
3. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press (1976)
4. Baron-Cohen, S.: Mindblindness: An Essay on Autism and Theory of Mind. MIT Press, Cambridge, MA, USA (1997)
5. Bavelas, J., Coates, L., Johnson, T.: Listener Responses as a Collaborative Process: The Role of Gaze. Communication **52**(3), 566–580 (2002)
6. Bee, N., André, E., Vogt, T., Gebhard, P.: Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues, chap. The Use of Affective and Attentive Cues in an Empathic Computer-Based Companion, pp. 131–142. John Benjamins (2010)
7. Behrooz, M., Rich, C., Sidner, C.: On the Sociability of a Game-Playing Agent: A Software Framework and Empirical Study. In: Intelligent Virtual Agents, IVA '14, pp. 40–53 (2014)
8. Bohus, D., Horvitz, E.: Multiparty Turn Taking in Situated Dialog: Study, Lessons, and Directions. In: SIGDIAL '11, pp. 98–109 (2011)
9. Carpenter, B., Pereira, F.: Computational Linguistics, vol. 19, chap. The Logic of Typed Feature Structures. Cambridge University Press (1992)
10. Clark, H.H.: Using Language. Cambridge University Press (1996)
11. Clark, H.H., Wilkes-Gibbs, D.: Referring as a Collaborative Process. Cognition **22**, 1–39 (1986)
12. Damian, I., Baur, T., Lugrin, B., Gebhard, P., Mehlmann, G., André, E.: Games are Better than Books: In-Situ Comparison of an Interactive Job Interview Game with Conventional Training. In: Artificial Intelligence in Education, AIED '15, pp. 84–94 (2015)
13. Doherty-Sneddon, G., Phelps, F.G.: Gaze Aversion: A Response to Cognitive or Social Difficulty? Memory and Cognition **33**(4), 727–733 (2005)
14. Duncan, S.: Some Signals and Rules for Taking Speaking Turns in Conversations. Personality and Social Psychology **23**(2), 283–292 (1972)
15. Gebhard, P., Mehlmann, G., Kipp, M.: Visual SceneMaker - A Tool for Authoring Interactive Virtual Characters. Multimodal User Interfaces **6**(1-2), 3–11 (2012)
16. Heerink, M., Kröse, B., Evers, V., Wielinga, B.: The Influence of Social Presence on Acceptance of a Companion Robot by Older People. Physical Agents **2**(2), 33–40 (2008)
17. Holroyd, A., Rich, C., Sidner, C.L., Ponsler, B.: Generating Connection Events for Human-Robot Collaboration. In: Robot and Human Interactive Communication, RO-MAN '11, pp. 241–246 (2011)
18. Kendon, A.: Some Functions of Gaze-Direction in Social Interaction. Acta Psychologica **26**(1), 22–63 (1967)
19. Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thrisson, K., Vilhjlmsson, H.: Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In: Intelligent Virtual Agents, IVA '06, pp. 205–217 (2006)
20. Lalanne, D., Nigay, L., Palanque, P., Robinson, P., Vanderdonckt, J., Ladry, J.F.: Fusion Engines for Multimodal Input: A Survey. In: ICMI '09, pp. 153–160 (2009)
21. Lee, J., Marsella, S., Traum, D., Gratch, J., Lance, B.: The Rickel Gaze Model: A Window on the Mind of a Virtual Human. In: IVA '07, pp. 296–303 (2007)
22. Mehlmann, G., André, E.: Modeling Multimodal Integration with Event Logic Charts. In: Multimodal Interaction, ICMI '12, pp. 125–132 (2012)
23. Mehlmann, G., Janowski, K., Baur, T., Markus Häring, E.A., Gebhard, P.: Exploring a Model of Gaze for Grounding in HRI. In: Multimodal Interaction, ICMI '14, pp. 247–254 (2014)
24. Mundy, P., Newell, L.: Attention, Joint Attention, and Social Cognition. Current Directions in Psychological Science **16**(5), 269–274 (2007)
25. Nielsen, G.: Studies in Self Confrontation. Munksgaard, Copenhagen, Denmark (1962)
26. Nooraei, B., Rich, C., Sidner, C.: A Real-Time Architecture for Embodied Conversational Agents: Beyond Turn-Taking. In: ACHI '14, pp. 381–388
27. Oviatt, S.: The Human-Computer Interaction Handbook, chap. Multimodal Interfaces. Lawrence Erlbaum, New Jersey, USA (2008)
28. Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N., Rich, C.: Explorations in Engagement for Humans and Robots. Artificial Intelligence **166**(1-2), 140–164 (2005)
29. Traum, D., Leuski, A., Roque, A., Gandhe, S., DeVault, D., Gerten, J., Robinson, S., Martinovski, B.: Natural Language Dialogue Architectures for Tactical Questioning Characters. In: Army Science Conference (2008)
30. Yngve, V.H.: On Getting a Word in Edgewise. Meeting of the Chicago Linguistic Society pp. 657–677 (1970)

Gregor Mehlmann is researcher at the HCM lab. Before that he was employed at the DFKI. He received his Bachelor and Master degrees in computer science at the Saarland University. He is a lead developer of the VSM modeling software.

Kathrin Janowski is researcher at the HCM lab. She received her Bachelor and Master degrees in computer science and multimedia at the Augsburg University. Her research focuses on interactive social and empathic robots and virtual agents.

Elisabeth André is full professor leading the HCM lab. Before that she was principal researcher at the DFKI. She is a highly respected expert for intelligent user interfaces, interactive virtual agents as well as social and affective computing.