Designing Appropriate Feedback for Virtual Agents and Robots

Manja Lohse¹ and Herwin van Welbergen²

Abstract—The virtual agents’ and the social robots’ communities face similar challenges when designing appropriate feedback behaviors. The paper points out some of these challenges, namely developing behaviors for various embodiments, integration and behavior generation, synchronization within systems, and coordination in groups of systems and users. We describe some (preliminary) solutions to these problems. Based on the remaining challenges we discuss future research directions that will allow the fields to profit from each other and to jointly make progress towards their aim of developing systems for social interaction with humans.

I. INTRODUCTION
Designing feedback for advanced interfaces such as social robots and virtual agents is a multi-disciplinary effort, requiring expertise in many research areas, including computer animation, perception, cognitive modeling, emotions and personality, natural language processing, speech recognition, speech synthesis, and nonverbal communication. However, research in virtual agents and human-robot interaction has so far not necessarily been strongly linked. Each field has developed its own methods and systems. At the same time both fields draw on the same insights from human social research [1]. Moreover, they aim at developing systems for social interaction with humans that successfully communicate their internal states using various modalities. This is particularly challenging because agents often still lack human-like capabilities and, thus, the interaction is asymmetric [2]. Moreover, previous research has shown that the appropriateness of agents’ feedback is influenced by situational constraints, i.e., in task-oriented interaction the user needs very concrete knowledge about the system’s internal states and abilities as compared to conversations that are mere social exchanges of ideas [2]. Given this, the fields of human-robot interaction and virtual agents face interrelated challenges and we should strive to share solutions and insights gained while working on these challenges.

The paper discusses four challenges that both fields face: developing behaviors for various embodiments, integration and behavior generation, synchronization within systems, and coordination in groups of systems and users. All these challenges are discussed in connection to behavior generation because this is central to our research and in the focus of the workshop. Also some (preliminary) solutions to the challenges from our own work and other researchers in the fields are presented. The paper concludes with an outlook on our research aims that address some of the challenges that the paper points out.

II. CHALLENGES AND STATE OF THE ART
In the following, we summarize some challenges that we encountered when starting to link our own work on the generation of feedback behavior for virtual agents and robots. Even though we divided the challenges into sections, there is quite some overlap between them and the respective connections are pointed out in the paper.

A. Developing Behaviors for Various Embodiments
The first challenge is to develop behaviors that are reusable on various embodiments. One idea related to this is the question of how human-like the systems should be in order to raise the right expectations in the users [3] and to have adequate ways of communicating their internal states to them. Thus, each system needs appropriate repertoires of behaviors and expressions that fit the respective embodiment. For effective system design it would be very useful if these repertoires could be translated for different systems such that behaviors can be evaluated on various platforms and standard behaviors become available for reuse.

We developed own approaches to this problem. Our AsapRealizer [4] has specifically been designed to transfer behavior (e.g., synchronized speech, gesture, facial expression) specified in the Behavior Markup Language (BML, see also Section II-B) on different embodiments. Currently, AsapRealizer is used to steer a virtual 3D agent, a cartoon character, a NAO robot¹, the Flobi robotic head [5] and the Nabaztag robot rabbit². Thus far we have ignored the more limited expressivity of the robots and we directly map BML behaviors that are meant to steer a virtual human onto more or less equivalent robot behavior (see Figure 1). BML behaviors specify behavioral signals in a relatively abstract manner (for example using the text to be spoken for speech, or Ekman’s action units for facial expressions).

The Bonsai framework [6], developed at Bielefeld University, provides reuse of behaviors on different platforms by implementing them in so-called skills. Skills are state-based deployments of sensors and actuators and enable the robot to complete certain tasks, e.g., to follow a person or to learn the name of an object. So far Bonsai has been implemented in the robots BIRON [7] and NAO. The approach taken in Bonsai is complementary to that in AsapRealizer in that it allows the

¹http://www.aldebaran-robotics.com/en/
²http://www.nabaztag.com
Fig. 1. FACS 1 left (inner eyebrow raise), implemented on a virtual character using mesh deformation (left), the FLOBI robot by rotating the eyebrow motor counter-clockwise (middle) and on the NAO robot using the LEDs on the right eye (right).

elegant composition of higher level skills out of lower level skills, in providing sensor-based skills and in providing skills that combine sensing and acting. However, unlike the BML-based behaviors of AsapRealizer, Bonsai provides limited functionality for the synchronization of multiple skills which is further discussed in Section II-C.

B. Integration and Behavior Generation

Using the AsapRealizer and Bonsai on the different systems leads us to the next challenge which is integration. As has been mentioned above, designing feedback for virtual agents and social robots are interdisciplinary endeavors. Researchers have realized that ‘the scope of building a complete virtual human is too vast for any one research group’ [8]. Modular architectures and interface standards enable researchers in different areas to reuse each other’s work and thus allow easier collaboration between researchers in different research groups [9]. In this context, the SAIBA initiative proposes an architecture for virtual agents [10] that provides such a modular design. This architecture (Figure 2) features a modular ‘planning pipeline’ for real-time multimodal motor behavior of virtual agents, with standardized interfaces (using representation languages) between the modules in the pipeline. The SAIBA Intent Planner module generates a plan representation on the functional level, specified in the Functional Markup Language (FML). FML will represent what a virtual human wants to achieve: its intentions, goals and plans [11]. The exact syntactical representation for this is still under discussion. Heylen et al. [11] indicate that (among other things) context, communicative actions, content, mental state and social-relational goals could be elements in FML. The SAIBA Behavior Planner generates a plan representation that is incrementally specified through blocks written in the Behavior Markup Language (BML). The Realizer executes behavior specified in BML onto a (virtual) agent. BML provides a general, realizer-independent description of multimodal behavior that can be used to control a virtual human. BML expressions (see Figure 3 for a short example) describe the occurrence of certain types of behavior (facial expressions, gestures, speech, and other types) as well the relative timing of the actions.

C. Synchronization within Systems

One main challenge that has been addressed with BML is synchronization among behaviors. Humans’ modalities are mostly well synchronized, e.g., human communication makes use of gestures that are tightly coordinated with speech. If their synchronization is off, the meaning that is jointly conveyed by gestures and speech becomes harder to understand [12]. We found that, while virtual agent behavior can typically be executed without failure and the synchronization constraints are met precisely, when executing robot behavior, one needs to take the possibility for execution failure and asynchrony into account. Synchronization of gesture, speech, and other modalities is a challenging task for social robots, since the exact timing of robotic gesture can typically not be predicted very precisely beforehand by standard robot software [13], [14]. This issue could, to some extent, be alleviated by more precise prediction models [13].

Since human modality synchronization is not always without trouble either, believable robots could make use of human-like strategies to repair synchrony in addition to better prediction strategies. E.g., humans can make use of hold phases in gesture or pauses in speech to maintain synchrony [15]. Salem [14] provides a robotic implementation of this synchronization strategy. In addition to the use of hold phases and pauses, humans make use of continuous micro-adaptations in their speech and gesture timing to maintain synchrony [16]. Recent work in flexible and adaptive Text-To-Speech systems (like INPRO_iSS [17]) and flexible and adaptive behavior planning [4] allow us to implement such adaptations of ongoing speech and motion on robots as well. To what extent these adaptations may be applied while retaining believability and whether such adaptations result in robotic behavior that is evaluated as being more believable than the use of pauses and hold phases is an open research
and the behavior sequence that is generally only available
it requires knowledge on the semantics of the constraints
is not something a realizer can answer on its own, since
lost time. The decision which of these possibilities to take
Finally, following motions could be sped up to make up
simply delaying everything that follows could make sense.
sequence meaningless, and it must be aborted. In other cases,
or overrun might be an error that renders the whole following
arises while the robot executes the behavior. For example, an
virtual agents share several research challenges with respect
incremental (dialog) processing: fluent interaction requires for example that agents are able
to deal with information increments that are smaller than
the full sentences that are typically used as information
increments in text-to-speech and speech recognition systems.
Being able to process and act upon information in such
smaller increments enables social agents to exhibit interper-
sonal coordination strategies such as backchannel feedback
and smooth turn taking. The IU-model [19] is a concep-
tual framework for specifying architectures for incremental
processing (of both input and output) in speech-only dialog
systems. Several systems have recently been implemented
using the IU-model. To allow one to use the IU-model for
the design of virtual agents or robots, the main challenge is
to generalize it to provide mechanisms for multimodal fusion
and fission of input and output.
In the robotic field, an architecture designed explicitly for
fluent interaction with robots has been proposed by Hoffman
and Breazeal [20]. Their cognitive architecture enables a
robot to anticipate the actions it should take, given the
task and user interaction history. Anticipation is fed into
the system as a top down bias of the perception process,
allowing it to select actions more rapidly (e.g., sometimes
even without requiring the user to ask for them).

D. Coordination in Groups of Systems and Users

Human interactions are highly dynamic and responsive.
Therefore, also agents must be capable of fluent incremental
behavior generation and perception. The agent’s behavior
must be adapted on-the-fly to the behavior of the interlocutor,
to achieve natural interpersonal coordination. AsapRealizer
[4] was designed as a BML realizer that specifically satis-
fies these requirements for behavior generation for virtual
humans.

To achieve a more natural dialog with and between social
agents, they also require incremental (dialog) processing:
fluent interaction requires for example that agents are able
to deal with information increments that are smaller than
the full sentences that are typically used as information
increments in text-to-speech and speech recognition systems.
Being able to process and act upon information in such
smaller increments enables social agents to exhibit interper-
sonal coordination strategies such as backchannel feedback
and smooth turn taking. The IU-model [19] is a concep-
tual framework for specifying architectures for incremental
processing (of both input and output) in speech-only dialog
systems. Several systems have recently been implemented
using the IU-model. To allow one to use the IU-model for
the design of virtual agents or robots, the main challenge is
to generalize it to provide mechanisms for multimodal fusion
and fission of input and output.
In the robotic field, an architecture designed explicitly for
fluent interaction with robots has been proposed by Hoffman
and Breazeal [20]. Their cognitive architecture enables a
robot to anticipate the actions it should take, given the
task and user interaction history. Anticipation is fed into
the system as a top down bias of the perception process,
allowing it to select actions more rapidly (e.g., sometimes
even without requiring the user to ask for them).

III. RESEARCH DIRECTIONS

We have shown that the fields of social robotics and
virtual agents share several research challenges with respect
to the design of appropriate system feedback. Some of
these challenges are addressed by researchers already and
we have discussed building blocks that may contribute to
their solution. However, various open problems remain to be
addressed in future work. An overview of our own future
directions is given in Table I. We summed them up in the
following four main points:
1) Both the Bonsai robotic framework and the AsapRealizer for virtual humans have contributed to enabling developers to reuse the same set of skills on different embodiments. A future challenge is to identify what skills can be shared between robots/virtual agents and what skills are best expressed by behavior that is specifically tailored for a specific embodiment.

2) The SAIBA architecture—and specifically the BML and BML Realizers—has allowed the use of standardized architecture elements for virtual humans. BML has shown to be useful for robotics, and the robotic community has recently become involved in the development of the standard. Robot behavior is, in general, more error-prone than virtual human behavior. Thus, to generalize the BML specification for use with robots, one of the major challenges is to enhance BML with specification mechanisms for failure detection, repair, and the generation of appropriate feedback. Furthermore, to enable BML realizers that are currently used to steer virtual humans to steer robots, they should be enhanced to handle such specification mechanisms.

3) Robots can make use of several modalities to express their behavior (e.g. speech, gesture, gaze, facial expression). The synchronization between such modalities can be essential for the robot’s interaction partner to rapidly understand the robot’s intention. Like humans, robots cannot always achieve intra-modal synchrony. We therefore propose that robots are endowed with human-like strategies to repair their synchrony. AsapRealizer’s flexible behavior adaptation mechanisms and the INPRO\_ISS flexible TTS system could be used as building blocks for such strategies.

4) Previous research has indicated that endowing robots and virtual agents with abilities to allow interactional coordination can enhance the perceived fluency of the interaction, the rapport between robot/virtual agent and human, the perceived humanlikeness of the agent, etc. However, how exactly (e.g. on what modalities, to what extend) robot behavior should be employed to achieve these positive effects and how they contribute to the quality of the interaction is an open research question. We aim to provide subjective and objective metrics to measure the quality of the interaction with robots and virtual agents. These measures will then allow us to do experiments in which we measure and compare the contribution of different coordination strategies/embodiments/etc. to interaction quality. To allow a robot or virtual agent to coordinate smoothly with a human, it needs to be able to predict and anticipate the behavior of its interlocutor. Such predictions could partly come from interaction history with the user on the same task [20]. Anticipation requires that the agent is able to continuously adapt its ongoing behavior. Functionality for this is provided in the AsapRealizer. Another requirement for smooth interaction is the ability to incrementally process input and output. Ymir and ACE have provided implementations for virtual agents that are capable of doing this; the IU model provides a general architecture framework for doing this in dialogue systems. Our ongoing work on the articulated sociable agents platform (ASAP) aims at bringing the combination of all these features required for interactional coordination together in a single architecture framework.

Addressing all these research questions will help in generating readable feedback for different platforms by integrating modalities in an appropriate way. Moreover, measures will be identified that enable users to express their evaluation of how readable and appropriate the system behavior is. We are currently setting up a collaborative effort with researchers with backgrounds in control engineering (robotics), applied artificial intelligence, human-machine interaction, psychology, and computational linguistics to tackle these challenges.

**References**


