All Together Now

Introducing the Virtual Human Toolkit

Arno Hartholt, David Traum, Stacy Marsella, Ari Shapiro, Giota Stratou, Anton Leuski, Louis-Philippe Morency, Jonathan Gratch

Institute for Creative Technologies
University of Southern California
12015 Waterfront Drive
Playa Vista, CA 90094, USA
{hartholt, traum, marsella, shapiro, stratou, leuski, morency, gratch}@ict.usc.edu

Abstract. While virtual humans are proven tools for training, education and research, they are far from realizing their full potential. Advances are needed in individual capabilities, such as character animation and speech synthesis, but perhaps more importantly, fundamental questions remain as to how best to integrate these capabilities into a single framework that allows us to efficiently create characters that can engage users in meaningful and realistic social interactions. This integration requires in-depth, inter-disciplinary understanding few individuals, or even teams of individuals, possess. We help address this challenge by introducing the ICT Virtual Human Toolkit[1], which offers a flexible framework for exploring a variety of different types of virtual human systems, from virtual listeners and question-answering characters to virtual role-players. We show that due to its modularity, the Toolkit allows researchers to mix and match provided capabilities with their own, lowering the barrier of entry to this multi-disciplinary research challenge.

Keywords. Virtual Humans, Embodied Conversational Agents, Architectures, Standards, Real-Time Integrated Systems, Toolkits, Audio-Visual Sensing, Natural Language Processing, Nonverbal Behavior

1 Introduction

Virtual humans, autonomous digital characters who interact verbally and nonverbally with users, can be powerful tools in a wide range of areas, including the teaching of interpersonal skills, cognitive science studies, military training, medical education, and entertainment. In fact, virtual humans have moved from the lab to actual deployment in a variety of fields in recent years, from virtual patients [1, 2] to pedagogical agents [3, 4] and military training [5, 6].

However, while virtual humans have advanced in both capability as well as applicability, they are still in their infancy. Realizing their full potential requires 1) compelling characters that can engage users in meaningful and realistic social interac-

[1] https://vhtoolkit.ict.usc.edu

tions, and 2) an ability to develop these characters effectively and efficiently. There are several challenges associated with these goals.

First, in order for virtual humans to be effective, they need to exhibit a range of capabilities, simulating those of real humans. This includes the ability to perceive human behavior (particularly communicative behavior aimed at the virtual humans), to process and understand this behavior, and to reason and produce appropriate (verbal and nonverbal) behaviors. Specialized knowledge and considerable resources may be needed to significantly advance any of these virtual human abilities.

Second, it is important that these abilities not work only in isolation, they also need to be integrated into a larger system (and further, into systems of systems), where they can inform, influence and strengthen each other. For example, the meaning of a spoken word like "yeah" might have a different meaning if it is accompanied by a head nod and a smile versus a head shake and frowning. This interdependence of functionalities can create obstacles for research, since individual researchers seeking to advance a specific topic (e.g., natural language understanding) might need to work in a large team with expertise in all other relevant aspects of the system, in order to solve the full problem. On the other hand, the integrated nature of virtual humans also creates novel research opportunities. Researchers can explore which integrated abilities are essential and desired for which type of interaction, what minimal and preferred dependencies exist between these abilities, how they affect each other, and how they can best leverage each other in order to achieve greater effectiveness.

Third, even with the appropriate knowledge and resources available, virtual humans can still be costly to develop. They are large, complex systems, often lacking specific frameworks or solid standards. Furthermore, certain principles may only be understood in a narrow context and can be difficult to generalize across multiple domains. This limits the ability to re-use knowledge and assets, often resulting in the need to start new characters or systems from scratch.

To address these challenges, we introduce the *ICT Virtual Human Toolkit,* which is designed to aid researchers with the creation of embodied conversational agents. The Virtual Human Toolkit enables further research by offering a collection of modules, tools, and libraries, as well as a framework and open architecture that integrates these components. The Toolkit provides a solid basis for the rapid development of new virtual humans, but also serves as an integrated research platform to enable context-sensitive research in any of the Virtual Human subfields, taking advantage of and examining the impact on other system modules. Rather than focusing on a specific type of agent, the Toolkit offers a flexible framework for exploring the vast space of different types of agent systems.

The Toolkit is based on over a decade of multi-disciplinary science and contains a mix of research and commercial technologies to offer full coverage of subareas including speech recognition, audio-visual sensing, natural language processing, dialogue management, nonverbal behavior generation & realization, text-to-speech and rendering. Example virtual humans are included to illustrate how components can work together and to enable the sharing of these assembled systems to the research community.

In this paper, we will show how the Virtual Human Toolkit provides an integrated platform for a variety of different virtual human architectures and how these architectures can be applied to a wide range of research efforts. Section 2 discusses the background of inter-disciplinary virtual human research. Section 3 explains the overall architecture and APIs on which the Toolkit is based, while Section 4 discusses the specific modules that are included in the Toolkit. In section 5 we explore how the Toolkit and related technologies have been used to create a range of different virtual human systems. Section 6 ends with a conclusion and future work.

2 Background

Reducing barriers to virtual human research requires shared tools and architectures and the Virtual Human Toolkit builds on several attempts to address this need. Efforts to develop and share individual capabilities include automated speech recognition [7], perception [8], task modeling [9], natural language generation [10], animation [9, 11, 12, 13, 14], and text-to-speech systems [15].

Some standardization efforts have attempted to consolidate these. SAIBA [16] specifies on a high level the generation of multimodal behavior. It aims to define an interface between the intent planning and behavior planning phases, called the Function Markup Language [17], and in between the behavior planning and behavior realization phases, called the Behavior Markup Language (BML) [18]. Several BML realizers are publically available, including LiteBody [9], GRETA [11], Elckerlyc [12], SmartBody [13], and EMBR [14]. While these realizers often depend on custom extensions to the BML standard, they have been shown to be compatible in real-time larger systems [19].

Further work has focused on including these capabilities and standards in larger frameworks. LiteBody and DTask form the basis for "Relational Agents" [9]. These agents aim to form long-term, social-emotional relationships with users. While these agents exhibit social verbal and nonverbal behavior, they are usually limited in their natural language processing and are often confined to 2D representations.

The Virtual People Factory [1] allows for the creation of virtual humans by domain experts themselves, rather than system experts. It is mostly used for creating virtual patients for medical and pharmacy education. Using a crowd sourcing approach focused on natural language interaction, new characters can be created more rapidly than through more traditional authoring methods.

GRETA [11] is a SAIBA compliant embodied conversational agent that focuses less on natural language interaction and more on affective nonverbal behavior generation and realization. In particular, it employs a model of complex facial generation.

The SEMAINE project [20] aims to integrate various research technologies, including some of the above, into creating a virtual listener. The emphasis is on perception and back-channeling rather than deep representations of dialogue.

Virtual Humans are not the only area of research in which these kinds of frameworks are desired. For example, the Robotic Operating System [21] defines structures that facilitate distributed development and sharing of robotic capabilities.

While these systems are both capable and successful, they typically focus on a limited subset of capabilities. In addition, few standards exist across capabilities, making efforts for further interoperability between them cumbersome.

3 Architecture, APIs and Virtual Human Capabilities

The Virtual Human Toolkit is the next step towards more integrated frameworks. It aims to offer a more comprehensive set of integrated capabilities than has been previously attempted, within a framework that allows for the rapid creation of new characters and systems. The Virtual Human Toolkit is an instantiation of a more general, modular Virtual Human Architecture, see Figure 1. This architecture defines at an abstract level the capabilities of a virtual human and how these capabilities interact. Not every virtual human system will include all the capabilities, and some will implement capabilities to a greater or lesser extent. The architecture also allows for multiple implementations of a certain capability and simple substitution of one implementation for another during run-time, facilitating the exploration of alternative models for realizing individual capabilities. Thus the general architecture can be specialized in many different ways, as we'll discuss in section 5. As presented in section 4, the Toolkit contains a set of modules providing at least one possible realization of each the capabilities.

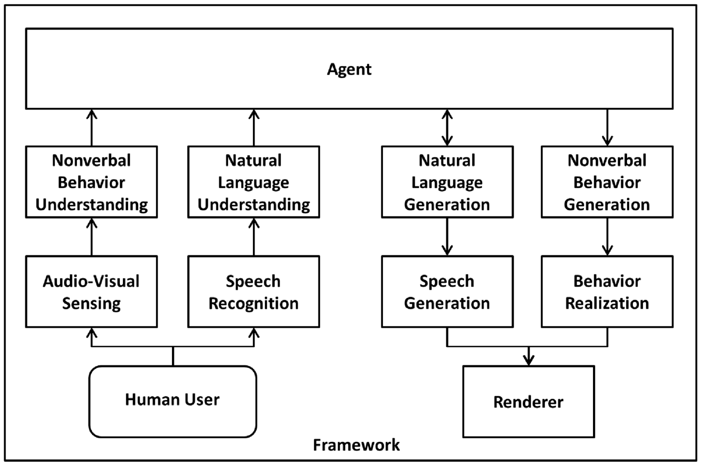
Agent
Nonverbal Behavior Understanding
Natural Language Understanding
Natural Language Generation
Nonverbal Behavior Generation
Audio-Visual Sensing
Speech Recognition
Speech Generation
Behavior Realization
Human User
Renderer
Framework

Fig. 1. The Virtual Human Architecture.

A virtual human system typically contains a subset of the capabilities depicted in Figure 1. A human user interacts with the system, which transforms the user's speech to a textual representation using speech recognition. This text is translated to a semantic representation through natural language understanding, which a dialogue manager within the agent can reason with. Audio-visual sensing relies on sensory input and has the ability to localize features and recognize specific expressions of nonverbal communication. Nonverbal behavior understanding combines information from different modalities in order to link certain observations through tracking and recognition to higher-level nonverbal communicative behaviors. Based on these inputs and internal state, the agent will create a communicative intent. This intent is further fleshed out both verbally and nonverbally through natural language generation and nonverbal behavior generation processes. Speech can be generated on the fly (e.g. text-to-speech) or be pre-recorded audio. Behavior realization will synchronize all behaviors (speech, gestures, lip synching, facial expressions, etc.) for a renderer to show.

API	Producer	Consumer	Description
vrSpeech	Speech Recognition	Natural Language Understanding	Text of partial or finalized user speech; also prosody, emotion.
vrNLU	Natural Language Understanding	Agent	Semantic representation of user's verbal input.
vrPerception	Nonverbal Behavior Understanding	Agent	Representation of user's nonverbal behavior in PML [34].
vrGenerate	Agent	Natural Language Generation	Communicative intent.
vrGeneration	Natural Language Generation	Agent	Surface text.
vrExpress	Agent	Nonverbal Behavior Generation	Communicative and functional intent, in FML and BML.
vrSpeak	Nonverbal Behavior Generation	Behavior Realization	Instructions of desired behavior in BML.

Table 1. Main elements of the Virtual Human API.

Capabilities are realized through specific modules, most of which communicate with each other through a custom messaging system called VHMsg, build on top of ActiveMQ [39]. Messages are typically broadcasted, although there are intended consumers. Libraries have been built for several languages, including Java, C++, C#, Lisp and TCL, so that developers have a wide latitude in developing new modules that can communicate with the rest of the system. There is a standard set of message types used by existing modules (see Table 1), and it is very easy to create new message types, as needed for new kinds of communications. Systems typically include a customizable launcher application, which allows launching, monitoring, and quitting of individual modules. The launcher can be configured with different sets of modules (to support different systems), or with different versions, to support individual customization and experimentation. There is also a logger that keeps track of all VHMsg traffic, and allows customized reporting. These three components together allow for easy reconfiguration and experimentation with specific system architectures.

4 Toolkit Provided Modules

The Toolkit contains a variety of specific modules that collectively cover the areas of speech recognition, audio-visual sensing, natural language processing, dialogue management, and nonverbal behavior generation & realization. See Figure 2 for how these are assembled within the context of the Toolkit. The modules are discussed in more detail below. Together with available authoring and debugging tools they can be used immediately to create new so-called question-answering characters (section 5.1). The Toolkit contains two examples of such characters, called Brad and Rachel (see Figure 3), as well as several other characters. The Toolkit's main target platform is Windows, with limited support for MacOS and Linux. Further support for these platforms, as well as Android and iOS is in development.

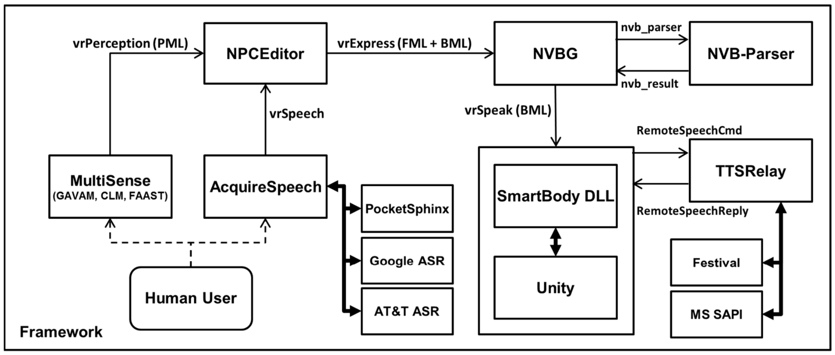

Fig. 2. Virtual Human Toolkit Architecture; regular lines are messages, bold lines direct links.

Fig. 3. Rachel (left) and Brad (right).

4.1 MultiSense

MultiSense provides the capabilities of both audio-visual sensing and nonverbal behavior understanding, as illustrated in Figure 1. The output messages are broadcast using the Perception Markup Language (PML) [34]. MultiSense is a multimodal sensing framework which is created as a platform to integrate and fuse sensor technologies and develop probabilistic models for human behavior recognition. MultiSense tracks and analyzes facial expressions, body posture, acoustic features, linguistic patterns and higher-level behavior descriptors (e.g. attention, fidgeting). MultiSense can provide quantitative information about the human's behavior that enhances training and diagnostic scenarios (e.g., public speaking [51] or psychological distress assessment [34] and section 5.3, below). MultiSense is designed with a modular approach, including synchronized capture of multiple modalities such as audio and depth image via Microsoft Kinect sensor and RGB video via webcam device. MultiSense includes updated components from the Social Signal Interpretation framework (SSI) [30] but also integrates multiple tracking technologies including CLM-Z FaceTracker [31] for facial tracking (66 facial feature points), GAVAM HeadTracker [32] for 3D head position and orientation, skeleton tracking by Microsoft Kinect SDK [52] and FAAST [33] for skeleton action coding. MultiSense utilizes a multithreading architecture enabling these different technologies to run in parallel and in real-time. Moreover, the framework's synchronization schemes allow for inter-module cooperation and information fusion. By fusing the different tracker results one can create a multimodal feature set that can be used to infer higher level information on perceived human behavioral state such as attentiveness, agitation and agreement.

4.2 NPCEditor

In the current distribution of the Toolkit, natural language understanding and agent functions such as dialogue management and output decisions are handled by the *NPCEditor* [22]. At the core of the NPCEditor is a statistical text classification algorithm that selects the character's responses based on the user's utterances. A character designer specifies a set of responses and a set of sample utterances that should produce each response. The algorithm analyzes the text of the sample utterances and of the responses and creates a mathematical description of the "translation relationship" that defines how the content of an utterance is mapped to the content of a response. When NPCEditor receives a new (possibly unseen) utterance, it uses this translation information to build a statistical language model of what it believes to be the best response. It then compares this representation to every stored response and returns the best match. The details of the algorithm can be found in [23]. We have shown the algorithm to be effective for constructing limited domain conversational virtual characters [22]. The NPCEditor also contains a dialogue manager that specifies how to use the classifier results, as well as tracking user-defined aspects of dialogue context. There are several provided dialogue managers, as well as a scripting ability for users to create their own dialogue managers.

4.3 NVBG

The Toolkit uses the *NonVerbal Behavior Generator* (NVBG) [24] to plan the character's nonverbal responses. The NPCEditor invokes character speech by sending the surface text to NVBG. In addition, it can receive optional functional markup that indicates the communicative function of the dialogue and the emotions/attitudes of the speaker. NVBG analyzes the text and functional markup to propose nonverbal behaviors. The resulting behavior set is communicated as BML [18]. Note that while conceptually the NVBG receives pure FML [17], in practice, the need for the surface text requires the addition of BML. NVBG is a rule-based behavior planner designed to operate flexibly with the information it is provided: it generates behaviors given the information about the agent's mental processes and communicative intent, but in the absence of such information infers communicative functions from a surface text analysis, using a parser. The rules within NVBG determine which nonverbal behaviors should be generated in a given context and were crafted using psychological research on nonverbal behaviors as well as analyses of annotated corpora of human nonverbal behaviors.

4.4 SmartBody

Behavior realization is achieved by *SmartBody*, [13]. It realizes behavior requests as a suitable animation sequence that conforms to the constraints of a BML request, such as synchronizing the emphasis phase of a gesture with a particular word of speech. The BML specification is heavily focused on synchronizing behaviors for a conversational agent: speech, gaze, gesturing and head movements. SmartBody uses additional BML parameters to control other aspects of character motion, such as locomotion and object manipulation [25], as well as parameters to tune the performance of other behaviors, such as the speed of the joints that a character will use to affix their gaze on another object. The generation of animation for a virtual character requires complex interaction between various body parts in order to adhere to the animation constraints and achieve a satisfactory performance. SmartBody uses various methods to handle animation constraints, both using procedurally-generated control mechanisms [26], as well as by using motion data either captured or hand-generated by digital artists. For example, gazing uses an inverse-kinematics based approach on which additional motion is then layered [27]. Locomotion and reaching use large amounts of motion data to handle full-body control that is difficult to do procedurally [28]. In addition, SmartBody uses a retargeting system that allows the transfer of groups of data designed for one character onto another [29].

4.5 Other Modules and Tools

Users can interact with characters by typing in text and by using Automated Speech Recognition (ASR). There is a module called *AcquireSpeech* (see Figure 2), which connects microphones and ASRs to the Toolkit with a common API, so other modules do not have to be adapted individually. There are interfaces for a number of

3rd party ASRs (including PocketSphinx [7], Google ASR and AT&T Watson). The main rendering platform of the Toolkit is Unity [35], a proprietary game engine which offers a free version. The Toolkit contains a variety of character art assets and backgrounds. Ogre [36] is included as an open source game engine example. Text-to-speech is provided by a single *TTSRelay* module that can interface to several text-to-speech engines, including Festival [15], CereVoice [37] and MS SAPI [38]. A variety of tools help with authoring, developing, configuring and debugging virtual human characters and systems. In particular, the *Machinima Maker* allows authors to create cut-scenes in Unity. An integrated authoring tool, called *VHBuilder*, abstracts some of the more advanced capabilities of the NPCEditor and NVBG, offering novices a simplified way to quickly author basic virtual humans.

5 Creating New Virtual Human Systems

Due to its modularity, the Virtual Human Toolkit is both configurable and extendable. Large scale permutations of module combinations and their individual configurations can be stored as system profiles, with smaller variations saved as user profiles. Researchers can mix and match existing Toolkit modules with their own, when adhering to the defined messaging API's. Given this flexibility, the Virtual Human Toolkit offers broad support for a range of virtual human systems. This enables researchers to explore and develop best practices for the research, design and development of virtual humans. In the remainder of this section we will explore several types of virtual humans and how the Toolkit can support some of these systems.

5.1 Question-Answering Characters

A common type of agent is the *question-answering character* [22]. Such agents offer a user-driven, interview-type style of conversation where a single user question is answered with a single character response. There is little to no maintenance of dialogue state nor are there advanced turn-taking strategies. These systems focus on providing the user with particular information within a given domain, often leading to a fixed set of pre-defined responses from one or more characters. The Toolkit supports these types of characters through the NPCEditor, NVBG and SmartBody modules. Several related systems [40, 41, 42] have been deployed and evaluated.

5.2 Virtual Listeners

Virtual listeners are agents that aim to simulate human listening behavior, in particular verbal and nonverbal back-channeling. These systems are typically user-driven, one-to-one and face-to-face. Input is commonly nonverbal only, i.e. a user's head movements, length and prosody of speech, etc. The agent itself is often a rule-based system, matching certain input patterns to pre-defined character behavior or generated behavior (e.g. mirroring of head movements, nodding). Example virtual listeners are Rapport [43] and SEMAINE [20], and a public speaking trainer [51]. The

Rapport system is included in the Virtual Human Toolkit, using a predecessor of GAVAM (now part of MultiSense) and a custom rule selector. It is based on psycholinguistic theory and was designed to create a sense of rapport between a human speaker and virtual human listener. It has been used in many studies to gather evidence that it increases speaker fluency and engagement. The Rapport system exemplifies how our approach enables both the re-use of existing knowledge and technologies as well as the sharing of results with the larger research community.

5.3 Virtual Interviewers

With *virtual interviewers*, the focus of the conversation lies on gathering information from or assessing the human user. Interactions are driven by the agent or are mixed-initiative. Shifting some of the conversational burden to the agent requires increased dialogue management and natural language understanding capabilities. Since the 'interview' is executed within a known domain, agent responses can be crafted up-front. An example of a virtual interviewer is the SimCoach system, a web-based guide that helps navigate healthcare related resources [44]. Its dialogue manager, FLoReS [45], allows for the creation of forward looking, reward seeking dialogues. An initial integration of FLoReS with the Toolkit has been completed and will be released shortly. An example of a hybrid virtual interviewer / listener is SimSensei, a virtual human platform to aid in recognition of psychological distress [35]. SimSensei enables an engaging face-to-face interaction where the virtual human automatically reacts to the perceived user state and intent, through its own speech and gestures. From the humans' signals and behaviors, indicators of psychological distress are inferred to inform a healthcare provider or the virtual human. SimSensei is in active development and advances several modules, including MultiSense, FLoReS, Cerebella (the successor to NVBG) and SmartBody.

5.4 Virtual Role-Players

Virtual humans can be particularly effective at facilitating interactive dramas or training scenarios in which a user (or player) must interact with other characters. They have advantages over human role-players due to their inherent consistency, 24-7 availability and ability to portray elements that are difficult or impossible for real humans (e.g. certain wounds, effects of a stroke, etc.). Examples include the INOTS [6] and ELITE training systems for the Navy and Army respectively, in which a single user practices interpersonal skills with a virtual human role-player as part of a larger class. It combines the NPCEditor and SmartBody with a branching storyline and intelligent tutor. Gunslinger [46] is a mixed-reality, story-driven experience, where a single participant can interact verbally and nonverbally with multiple virtual characters that are imbedded in a physical saloon. Together with the NVBG and SmartBody, it uses a version of the NPCEditor which has been extended to incorporate the notion of hierarchical interaction domains, comparable to a state machine. In addition, it receives perception input which is treated as an additional token in the statistical analysis.

5.5 Virtual Confederates

Finally, virtual humans are gaining interest as a methodological tool for studying human cognition, including the use of *virtual confederates*. Virtual humans not only simulate the cognitive abilities of people, but also many of the embodied and social aspects of human behavior more traditionally studied in fields outside of cognitive science. By integrating multiple cognitive capabilities to support real-time interactions with people, virtual humans create a unique and challenging environment within which to develop and validate cognitive theories [48, 49, 50].

6 Conclusions and Future Work

We have shown how the Virtual Human Toolkit helps to address several challenges in virtual human research. First, a full virtual human system requires many different capabilities, and the Toolkit contains modules that cover audio-visual sensing, nonverbal behavior understanding, speech recognition, natural language processing, nonverbal behavior generation & realization, text-to-speech and rendering. Second, individual capabilities need to be integrated into a larger framework, which the Toolkit offers in the form of a reference architecture and related APIs. As such, it provides a rich context in which to embed individual research efforts and delivers a flexible framework that enables the exploration of a wide range of different virtual human systems. Finally, it lowers the effort associated with creating virtual humans by providing a suite of modules and tools that facilitate the rapid development of new characters and by promoting re-use of assets. These reduce the required knowledge and resources to develop virtual humans, lowering the barrier of entry into further inter-disciplinary research. Empirical evidence suggests a computer literate user with no particular virtual human or computer science background can build a limited question-answering character in less than a day, with more advanced characters taking up to several weeks. We plan to more formally evaluate these efforts within the year. While creating new virtual human architectures and systems can be done by individuals, depending on the complexity this may require a small group of specialists with a computer science background.

The Toolkit and related technologies have already been used in several dozen research and applied projects at ICT. Since its release to the community, it has seen download requests from close to 400 individuals and has been used as a teaching tool in several classes.

While the Toolkit offers a comprehensive framework for virtual human research and development, it is not without its limitations. The provided NPCEditor component focuses on statistical text classification rather than deep understanding of natural language and dialogue management. This limits verbal interactions to mostly question-answering type conversations. We aim to address this by releasing both the hierarchical interaction domain plugin for the NPCEditor as well as the FLoReS dialogue manager shortly. In addition, the Toolkit is lacking certain abilities, in particular task driven behavior, emotion modeling and persistent memory. While these areas are of interest in our basic research, we feel they are currently less appropriate for inclusion with the more applied Toolkit.

Our current focus is on expanding the capabilities of the Toolkit. In addition to including more advanced dialogue management, we aim to more tightly integrate MultiSense, to expand the available character library, and to increase the number of supported platforms, including full support for Mac OS, Android, iOS and the web.

Future work is aimed at addressing many of the lessons learned over the past decade. With Cerebella, the successor to the NVBG, we aim to expand our model of nonverbal behavior and provide a more powerful way to describe and generate this behavior, including the ability for a character to keep a consistent gesture space. This will require the implementation of gesture co-articulation as well as extension to BML. In addition, we aim to expand the role of nonverbal behavior understanding in order to provide a richer context to the agent. Furthermore, we will continue to investigate methods that support exploration of different configurations within the modular architecture, both at the level of modules themselves as well as the implementation of different gradations of capability compliance (e.g. per utterance speech recognition results vs. partial speech results vs. continuous speech). This requires a refinement of our current distributed messaging model, creating a balance between a structured and well-defined API on the one hand and a non-restrictive and expandable infrastructure that allows for experimentation on the other. Finally, we aim to more concretely define separate categories, or genres, of virtual humans. This allows for the creation of best practices, methodologies and supporting tools that enable more rapid development of virtual human systems.

Acknowledgements. We like to acknowledge all the hard work of our many colleagues at USC ICT without whom the Virtual Human Toolkit would not have been possible. The effort described here has been sponsored by the U.S. Army. Any opinions, content or information presented does not necessarily reflect the position or the policy of the United States Government; no official endorsement should be inferred.

7 References

1. Rossen B., Lok B., A crowdsourcing method to develop virtual human conversational agents, International Journal of HCS, pp. 301-319 (2012)
2. Bickmore, T.W., Bukhari L., Pfeiffer L., Paasche-Orlow M., Shanahan C., Hospital Buddy: A Persistent Emotional Support Companion Agent for Hospital Patients, Springer Berlin/Heidelberg, pp. 492-495 (2012)
3. D'Mello, S. K., & Graesser, A. C. AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. ACM Transactions on Interactive Intelligent Systems, Volume 2, Issue 4, Article 23 (2012)
4. Lane H.C., Noren,D., Auerbach D., Birch M., Swartout W., Intelligent Tutoring Goes to the Museum in the Big City: A Pedagogical Agent for ISE, Artificial Intelligence in Education, pp. 155-162 (2011)
5. Johnson, W.L., Valente, A., Tactical Language and Culture Training Systems: Using AI to Teach Foreign Languages and Cultures. pp. 72-83 (2009)

6. Campbell J., Core, M., Artstein R., Armstrong L., Hartholt A., Wilson C., Georgila K., Morbini F., Haynes E., Gomboc D., Birch M., Bobrow J., Lane H., Gerten J., Leuski A., Traum D., Trimmer M., DiNinni R., Bosack M., Jones T., Clark R., Yates K. Developing INOTS to support interpersonal skills practice. In Proceedings of the Thirty-second Annual IEEE Aerospace Conference, pp. 1-14, (2011)
7. Pocketsphinx: A Free, Real-Time Continuous Speech Recognition System for hand-Held Devices, vol. 1, pp. 185-188 (2006)
8. Littlewort G., Whitehill J., Wu T., Fasel I., Frank M., Movellan J., and Bartlett M., The Computer Expression Recognition Toolbox (CERT). Proc. IEEE International Conference on Automatic Face and Gesture Recognition (2011)
9. Bickmore, T.W., Schulman D., Shaw, G., DTask & LiteBody: Open Source, Standards-based Tools for Building Web-deployed Embodied Conversational Agents, Intelligent Virtual Agents, PP. 425-431 (2009)
10. Stone M., Specifying Generation of Referring Expressions by Example. AAAI Spring Symposium on NLG in Spoken and Written Dialogue, pp. 133-140 (2003)
11. I.Poggi, C.Pelachaud, F. de Rosis, V. Carofiglio, B. De Carolis, GRETA. A Believable Embodied Conversational Agent, Multimodal Intelligent Information Presentation, (2005)
12. Van Welbergen, H., Reidsma, D., Ruttkay, Z.M., Zwiers, J.: Elckerlyc: A BML realizer for continuous, multimodal interaction with a virtual human. Multimodal UI (2010)
13. Shapiro, A.: Building a Character Animation System. In: The Fourth International Conference on Motion in Games, Edinburgh, UK, November, 2011
14. Heloir, A. and Kipp, M.: A Realtime Engine for Interactive Embodied Agents, Intelligent Virtual Agents, pp. 393-404, 2009
15. Taylor, P., Black, A., Caley, R., The architecture of the Festival speech synthesis system. Third ESCA Workshop in Speech Synthesis, pp. 147–151. (1998)
16. http://www.mindmakers.org/projects/saiba/wiki
17. Heylen, D.K.J., Kopp, S., Marsella, S.C., Pelachaud, C., Vilhjálmsson, H.H. The Next Step towards a Function Markup Language. IVA (2008)
18. Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Th´orisson, K., Vilhj´almsson, H., Towards a Common Framework for Multimodal Generation: The Behavior Markup Language (LNAI), vol. 4133, pp. 205–217. (2006)
19. H. van Welbergen, Y. Xu, M. Thiebaux, W.W. Feng, J. Fu, D. Reidsma and A. Shapiro Demonstrating and Testing the BML Compliance of BML Realizers. IVA (2011)
20. Schröder, M. (2010). The SEMAINE API: Towards a standards-based framework for building emotion-oriented systems. Advances in Human-Machine Interaction (2010)
21. Quigley, M., Conley, K., Gerkey, B.P., Faust, J., Foote, T., Leibs, J., Wheeler, R., and Ng, A.Y., ROS: an open-source Robot Operating System, ICRA Open Source Software, (2009)
22. Leuski, A and Traum, D. NPCEditor: Creating virtual human dialogue using information retrieval techniques. AI Magazine, 32(2):42–56, (2011).
23. Leuski, A., Traum D.. A statistical approach for text processing in virtual humans. In Proceedings of the 26th Army Science Conference, Orlando, Florida, USA, December 2008.
24. Lee, J. and S. Marsella. Nonverbal Behavior Generator for Embodied Conversational Agents, IVA (2006)
25. Feng, A.W., Xu, Y, Shapiro, A. : An Example-Based Motion Synthesis Technique for Locomotion and Object Manipulation, SIGGRAPH (2012)
26. Thiebaux, M., Lance, B., Marsella, S.: Real-time Expressive Gaze for Virtual Humans, AAMAS, vol. 1, pp. 321-328 (2008)
27. Kallmann, M., Marsella, S.: Hierarchical Motion Controllers for Real-time Autonomous Virtual Humans, Intelligent Virtual Agents, pp. 253-265, (2005)

28. Feng, A.W., Huang, Y, Shapiro, A., An Analysis of Motion Blending Techniques, The Fifth International Conference on Motion in Games, Rennes, France, November 2012
29. Feng, A.W., Huang, Y, Xu, Y, Shapiro, A., Automating the Transfer of a Generic Set of Behaviors Onto a Virtual Character, Conference on Motion in Games, (2012)
30. J. Wagner, F. Lingenfelser, N. Bee, and E. Andre. Social signal interpretation (ssi). KI - Kuenstliche Intelligenz, 25:251–256, (2011)
31. T. Baltrusaitis, P. Robinson, and L.-P. Morency. 3D constrained local model for rigid and non-rigid facial tracking. IEEE Computer Vision and Pattern Recognition, June (2012)
32. L.-P. Morency, J. Whitehill, and J. Movellan. Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation. Automatic Face and Gesture Recognition, pp. 1–8, (2008)
33. E. Suma, B. Lange, A. Rizzo, D. Krum, and M. Bolas, "FAAST: The Flexible Action and Articulated Skeleton Toolkit," Proceedings of IEEE Virtual Reality, pp. 247-248, (2011).
34. Scherer, S., Marsella, S., Stratou G., Xu X., Morbini F., Egan A., Rizzo A., Morency, L.P., Perception Markup Language: Towards a Standardized Representation of Perceived Nonverbal Behaviors, IVA pp. 455–463 (2012)
35. http://unity3d.com/
36. http://www.ogre3d.org/
37. http://www.cereproc.com/products/sdk
38. http://msdn.microsoft.com/en-us/library/ee125663(v=vs.85).aspx
39. http://activemq.apache.org/
40. Leuski A, Kennedy B, Patel R, Traum DR. Asking questions to limited domain virtual characters: how good does speech recognition have to be? ASC (2006)
41. Artstein R, Gandhe S, Leuski A, Traum DR. Field Testing of an interactive question-answering character. ELRA, LREC (2008)
42. Swartout W, Traum DR, Artstein R, Noren D, Debevec P, Bronnenkant K, et al. Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides. IVA pp. 286–300 (2010)
43. Jonathan Gratch, Anna Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, R. J. van der Werf, Louis-Philippe Morency, Virtual Rapport, IVA (2006)
44. Rizzo A., Forbell E., Lange B., Buckwalter J.G., Williams J., Sagae K., Traum D., SimCoach: An Online Intelligent Virtual Agent System for Breaking Down Barriers to Care for Service Members and Veterans, Chapter in Healing War Trauma (2012)
45. Morbini F., DeVault D., Sagae K., Gerten J., Nazarian A., Traum D., FLoReS: A Forward Looking, Reward Seeking, Dialogue Manager, Spoken Dialog Systems (2012)
46. Hartholt A., Gratch J., Weiss L., Leuski A., Morency L.P., Marsella S, At the Virtual Frontier: Introducing Gunslinger, a Multi-Character, Mixed-Reality, Story-Driven Experience IVA pp. 500-501 (2009)
47. Kenny, P.G., Parsons, T.D., Gratch, J. Rizzo, A., Evaluation of Justina: A Virtual Patient with PTSD, Lecture Notes in Computer Science, pp. 394-408 (2008)
48. Traum DR, Marsella S, Gratch J, Lee J, Hartholt A. Multi-party, Multi-issue, Multi-strategy Negotiation for Multi-modal Virtual Agents. IVA pp. 117-130(2008)
49. Gratch J., Hartholt A., Dehghani M., Marsella S., Virtual Humans: A New Toolkit for Cognitive Science Research, CogSci (2013)
50. Khooshabeh P., McCall C., Gandhe S., Gratch J., Blascovich J., Does it matter if a computer jokes, Extended abstracts on Human Factors in Computer Systems, pp. 77-86 (2011)
51. Batrinca, L., Stratou, G., Morency, L.P., Scherer, S. Cicero - Towards a Multimodal Virtual Audience Platform for Public Speaking Training. IVA (2013)
52. http://www.microsoft.com/en-us/kinectforwindows/