VIRTUAL MUSEUM GUIDES DEMONSTRATION

William Swartout[1] David Traum[1] Ron Artstein[1] Dan Noren[2] Paul Debevec[1] Kerry Bronnenkant[2] Josh Williams[1] Anton Leuski[1] Shrikanth Narayanan[3] Diane Piepol[1] Chad Lane[1] Jacquelyn Morie[1] Priti Aggarwal[1] Matt Liewer[1] Jen-Yuan Chiang[1] Jillian Gerten[1] Selina Chu[3] Kyle White[3]

[1]USC Institute for Creative Technologies
[2]Museum of Science, Boston
[3]USC Signal Analysis and Interpretation Laboratory

The Virtual Museum Guides [1] are two virtual humans set in an exhibit at the Museum of Science, Boston, designed to promote interest in Science, Technology, Engineering and Mathematics (STEM). The primary audience is children between ages 7 to 14, in particular females and other groups under-represented in STEM.

The Guides are based on and extend the approach used in the SGT Star character [2] and the Gunslinger project [3]. To interact with the characters, an operator presses a push-to-talk button and speaks into a microphone. An audio acquisition client then sends audio to the automatic speech recognizer (ASR), which creates hypotheses of the words that were said, and then sends this text to the Language Understanding (LU) module. The ASR module uses the SONIC toolkit [4], with custom language and acoustic models. The LU module uses a statistical text classification algorithm to map the utterance text onto character responses. It selects a set of responses it believes to be appropriate to the utterance from a domain-specific library of scripted responses and passes them to the dialogue management (DM) module. The DM module uses that response set and the recent dialogue history to select one response for the characters to perform. The response is sometimes a sequence of utterances; in this case, the DM keeps a schedule of pending utterances, and sends them one at a time to the animation components, waiting for a callback signal before sending the next one. If the characters are interrupted by more speech from the operator before the schedule has completed, the DM can cancel the remaining sequence.

The LU/DM module pair uses the NPCEditor software [5]. The NPCEditor classification algorithm analyzes the text of the sample utterances and the text of the responses and creates a statistical model of the "translation relationship" that defines how the content of an input utterance determines the likely appropriateness of a response. Specifically, it learns how to compute a conditional likelihood of observing a particular word in a character's response given an operator's utterance [6]. When NPCEditor receives a new (possibly unseen) utterance, it uses this translation information to build a model of what it believes to be the best response for the utterance. The classifier then compares this representation to every stored response and returns the best match to the DM part of NPCEditor. In contrast, a traditional text classification approach would compare a new question to sample questions and then directly return the corresponding responses, ignoring the actual text of the response. We have observed that this "translation-based" classification approach significantly increases the effectiveness of the classifier for imperfect speech recognition [6]. NPCEditor has been fielded in a number of applications and has been shown to be successful in noisy classification tasks [5].

The Guides have a large but finite set of responses (currently about 400), so the characters might repeat themselves. One of the tasks of the DM is to match the classifier selection to the recent dialogue history and choose responses that have not been heard. The DM also handles cases when the classifier returns no responses. This happens when the operator asks a question for which the characters have no answer or speech is not understood by the ASR module. In that case, the classifier decides that none of the known answers is appropriate. The characters' database contains a number of responses that we call "off-topic." These responses range from prompts for repetition "Could you ask that again?" to utterances indicating that the characters do not know how to answer the questions "I really wish we had an answer for that."

The animation process is revised from that used by SGT Star and employs the Smartbody (SBM) behavior realization system [7] and a new sequencer module, as well as the Gamebryo animation engine. The sequencer module retrieves Behavior Markup Language (BML) [8] animation schedules for each of the utterances coming from the DM. Since BML as interpreted by SBM only animates a single agent, the sequencer schedule includes a number of synchronization points that are broadcast back to the sequencer. When the sequencer receives these callbacks it sends additional BML schedules to animate

This abstract is taken from Swartout et al. [1], "Ada and Grace: Toward realistic and engaging virtual museum guides," IVA 2010, LNAI 6356, section 3 (pp. 290–291) and Figure 1 (p. 286).

the other agent, so that the two characters can each react appropriately while the other is speaking. SBM uses several behavior controllers and blending to realize the specific combination of motion, and sends the resulting commands to the Gamebryo engine to generate the motion.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant 0813541. We thank the staff and volunteers of Cahners ComputerPlace for their support. We also thank Kim LeMasters, Creative Director of the ICT, for the suggestion to use twins, Ed Fast for software support, and Stacy Marsella and Andrew Marshall for help with the Smartbody system. Finally, we would like to thank Arlene de Strulle for her continued support and enthusiasm.

REFERENCES

[1] William Swartout, David Traum, Ron Artstein, Dan Noren, Paul Debevec, Kerry Bronnenkant, Josh Williams, Anton Leuski, Shrikanth Narayanan, Diane Piepol, Chad Lane, Jacquelyn Morie, Priti Aggarwal, Matt Liewer, Jen-Yuan Chiang, Jillian Gerten, Selina Chu, and Kyle White, "Ada and Grace: Toward realistic and engaging virtual museum guides," in *Intelligent Virtual Agents: 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20–22, 2010 Proceedings*, Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova, Eds., vol. 6356 of *Lecture Notes in Artificial Intelligence*, pp. 286–300. Springer, Heidelberg, September 2010.

[2] Ron Artstein, Sudeep Gandhe, Jillian Gerten, Anton Leuski, and David Traum, "Semi-formal evaluation of conversational characters," in *Languages: From Formal to Natural. Essays Dedicated to Nissim Francez on the Occasion of His 65th Birthday*, Orna Grumberg, Michael Kaminski, Shmuel Katz, and Shuly Wintner, Eds., vol. 5533 of *Lecture Notes in Computer Science*, pp. 22–35. Springer, Heidelberg, May 2009.

[3] Arno Hartholt, Jonathan Gratch, Lori Weiss, and The Gunslinger Team, "At the virtual frontier: Introducing Gunslinger, a multi-character, mixed-reality, story-driven experience," in *Intelligent Virtual Agents: 9th International Conference, IVA 2009, Amsterdam, The Netherlands, September 14–16, 2009 Proceedings*, Zsófia Ruttkay, Michael Kipp, Anton Nijholt, and Hannes Högni Vilhjálmsson, Eds., vol. 5733 of *Lecture Notes in Artificial Intelligence*, pp. 541–542. Springer, Heidelberg, September 2009.

[4] Bryan Pellom and Kadri Hacıoğlu, "Sonic: The University of Colorado continuous speech recognizer," Tech. Rep. TR-CSLR-2001-01, Center for Spoken Language Research, University of Colorado, Boulder, 2001.

[5] Anton Leuski and David Traum, "NPCEditor: A tool for building question-answering characters," in *Proceedings of The Seventh International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta, 2010, pp. 2463–2470.

[6] Anton Leuski and David Traum, "Practical language processing for virtual humans," in *Twenty-Second Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-10)*, Atlanta, GA, 2010.

[7] Marcus Thiebaux, Andrew N. Marshall, Stacy Marsella, and Marcelo Kallmann, "Smartbody: Behavior realization for embodied conversational agents," in *International Conference on Autonomous Agents and Multi-Agent Systems*, Estoril, Portugal, 2008, pp. 151–158.

[8] Hannes Vilhjálmsson, Nathan Cantelmo, Justine Cassell, Nicolas E. Chafai, Michael Kipp, Stefan Kopp, Maurizio Mancini, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Zsofi Ruttkay, Kristinn R. Thórisson, Herwin van Welbergen, and Rick J. van der Werf, "The behavior markup language: Recent developments and challenges," in *Intelligent Virtual Agents: 7th International Conference, IVA 2007, Paris, France, September 17–19, 2007 Proceedings*, Catherine Pelachaud, Jean-Claude Martin, Elizabeth André, Gérard Chollet, Kostas Karpouzis, and Danielle Pelé, Eds., vol. 4722 of *Lecture Notes in Artificial Intelligence*, pp. 99–111. Springer, Heidelberg, September 2007.

The Guides at the Museum of Science, Boston