Trackside DEIRA: A Dynamic Engaging Intelligent Reporter Agent (Demo Paper)

François L.A. Knoppel, Almer S. Tigelaar, Danny Oude Bos, Thijs Alofs, Zsofia Ruttkay
Human Media Interaction, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands
{f.l.a.knoppel, a.s.tigelaar, d.oudebos, t.alofs} @student.utwente.nl,
z.m.ruttkay@ewi.utwente.nl

ABSTRACT
DEIRA is a virtual agent commenting on virtual horse races in real time. DEIRA analyses the state of the race, acts on emotion and comments about the situation in a believable and engaging way, using synthesized speech and facial expressions. This paper shortly describes the features of this embodied conversational agent.

Categories and Subject Descriptors
I.2.7 [**Artificial Intelligence**]: Language Generation and Speech Synthesis; J.4 [**Computer Applications**]: Social and Behavioral Sciences; H.5.1 [**Multimedia Information Systems**]: Animations and Audio output; I.3.7 [**Three-Dimensional Graphics and Realism**]: Animation

General Terms
Algorithms, Performance, Design, Experimentation, Human Factors, Languages.

Keywords
Intelligent virtual agent, multimodal communication, emotion modeling, facial expressions, synthetic speech.

1. INTRODUCTION
Within the context of the GALA[1] contest, we have developed an embodied conversational agent (ECA) that provides commentary for virtual horse races. This specific challenge of producing such a horse race reporter was posed to the ECA community by GALA to provide a well-defined domain which at the same time is rich enough to place demands on both the cognitive and presentational skills of the reporter.

At the GALA contest, our agent was reporting real-time on a race script supplied on the spot. A second race reporter was present, which made comparison of the reactions and overall performance possible [2]. At this competition, Trackside DEIRA won both the public and jury awards[2].

[1] See http://hmi.ewi.utwente.nl/gala/racereporter

[2] See http://hmi.ewi.utwente.nl/gala/finalists_2007

Cite as: Trackside DEIRA (Demo Paper), Knoppel, F.L.A. et al., *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, 1681-1682.

2. FEATURES
During the race, information on the location and speed of each of the horses is sent from the HRSS to the Race Analysis Module (RAM) every second. The RAM uses a *rule-based analysis* to detect many different types of events, making it fast and easy to adjust.

The Mental Model Module (MMM) determines the *emotional impact* (defined in levels of excitement, surprise, amusement, and pity) of each event based on the event type and personality of the reporter. It also supplies a general *emotional state* which is a combined result of all events that have occurred taking into account a reduced influence over time and the assigned importance of the events.

After passing the MMM the events are stored in a *prioritized* Event Queue (EQ), based on an importance and importance decay factor as defined by the rules defined for the RAM. This ensures that the most important events are reported first. The EQ also decreases chances of repetition by lowering the importance of events that are similar to those recently uttered.

Using a generative context-free grammar supporting variables and conditionals, the Text Generation Module (TGM) constructs a set of potential utterances for each event it gets from the EQ. To prevent repetition, a history of utterances is maintained. The grammar provides *a rich vocabulary* to report each event, and is easy to expand or adjust for other purposes.

For *vocal expression*, the Speech Adaptation Module (SAM) determines at what speed, pitch and volume the sentences should be uttered based on the excitement of the reporter and the event itself.

The *facial expression* is handled by the Facial Animation Module (FAM) in two ways. Idle head movements (like saccadic eye movements and small head motions) are performed at fixed intervals. When reporting on an event the emotional state of the reporter is translated into a more conspicuous expression like smiling and frowning. A higher excitement leads to more pronounced movements for both the idle and event-based animations.

The face, voice, and emotional personality of the reporter are all defined in the *Personality* DataBase (PDB).

The Output Module (OM) feeds the text plus utterance characteristics and animation data to the external VisageLink[3]

[3] For more information concerning Visage: Technologies, visit: http://www.visagetechnologies.com

application to animate the head and instruct the Text-To-Speech (TTS) engine to generate speech, using Nuance's RealSpeak™ US English voice.

Lip synchronization is done by the TTS engine and Visage.

DEIRA is *highly reactive*, with a minimal response time of 30 ms, and one second on average as events are often held back until the reporter finishes commenting on a previous one.

During lengthy uneventful periods, *background events* are used to fill the quiet moments.

The highly *modular design* of the system and the separate declaration of the domain-specific knowledge make it easy to adapt it to different applications (e.g. RoboSoccer reporting [1], virtual storytelling) and user groups (e.g. language, age group). To achieve this, the proper event rules and parameters should be defined, plus the appropriate event-to-text grammar. Visualization could be made more suitable to the application by changing the face and voice settings in the PDB.

In its current state, the external Horse Race Simulation Software (HRSS) can be offered a scripted race to report on as well as let a race directly controlled by the user. The rich vocabulary of the system in combination with randomization means that even when using scripted races, *variety* in commentary is ensured between runs.

In addition to these features, because of the set of event generation rules and event-to-text grammar are stored in a human-readable format in simple text files, the output of the system can be altered significantly within a few minutes, displaying the *flexibility* of the system.

Figure 1 shows the entire system in action. At the top the VisageLink application shows the reporter. In the middle the HRSS is shown and below that, the Trackside DEIRA application displays both the generated sentences and emotional state.

3. CONCLUSIONS

DEIRA is a highly reactive virtual horse race reporter, capable of interpreting raw race observation data, conveying his emotional state and commenting on race events by speech accompanied by facial expressions in real-time. The system uses rule-based technology for event analysis and mental models, natural language processing and a third-party TTS engine and facial model for output synthesis.

We have received positive feedback from professional circles regarding the mental and communicational capabilities of DEIRA and the overall engaging experience. Evaluation results show that potential users are impressed by the capabilities of the system. They preferred having our virtual reporter commenting on a virtual race, as opposed to a race system without one, signifying a clear added value of our system. The same people were still critical in comparing the virtual reporter to a real one.

For more details on implementation and user evaluation results, please refer to our full article which is included in the AAMAS 2008 conference proceedings.

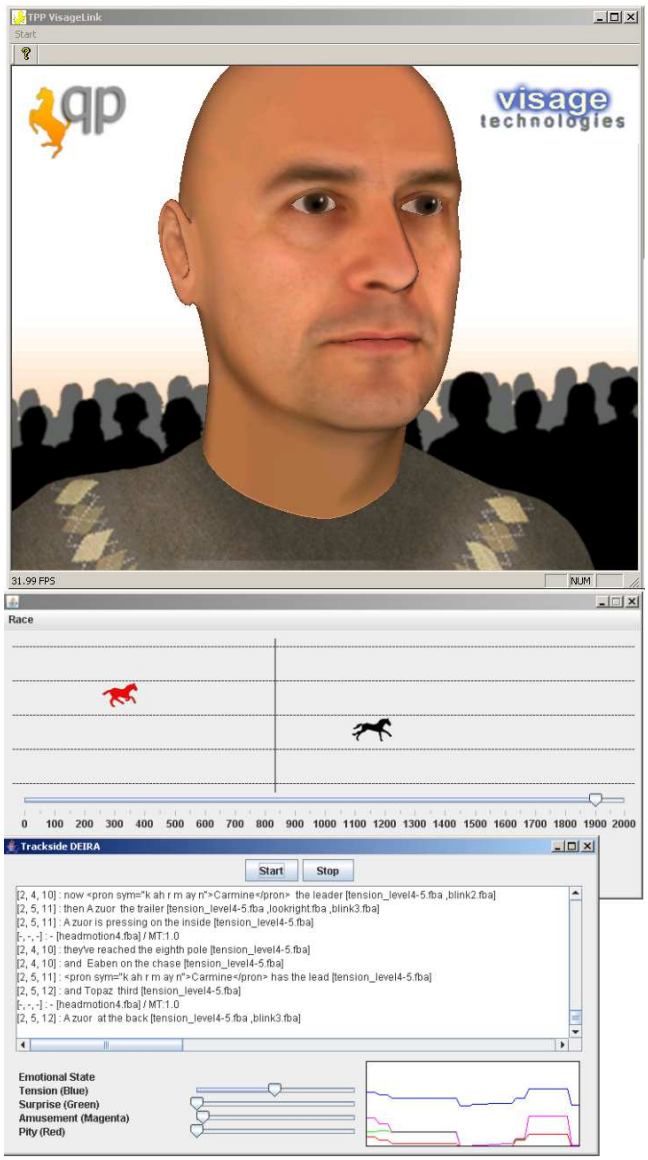

Figure 1. System screenshot

4. ACKNOWLEDGEMENTS

We would like to thank the GALA Organization, all those that were willing to participate in our enquiry and our friends and family. We furthermore want to thank Prof. Igor Pandzic for his cooperation in our use of Visage.

5. REFERENCES

[1] Andre, E., Binsted, K., Tanaka-Ishii, K., Luke, S., Herzog, G. and Rist, T. 2000. Three RoboCup simulation league commentary systems. AI Magazine 21 (1), 57-66.

[2] Strauss, M., Kipp, M. 2008. ERIC: A Generic Rule-based Framework for an Affective Embodied Commentary Agent. Submitted to AAMAS08.