

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

François L.A. Knoppel Almer S. Tigelaar Danny Oude Bos Thijs Alofs

Human Media Interaction, University of Twente, P.O.Box 217 7500AE Enschede

Abstract

DEIRA is an embodied agent with a highly modular design, supporting several domains such as real time virtual horse race commentary, robosoccer commentary and virtual storytelling. Domain-specific information is processed to have the agent act on emotion and produce a compelling report on the situation, using synthesized speech and facial expressions. This paper briefly describes the features of the agent.

1 Introduction

DEIRA was originally developed as an embodied agent that provides commentary for real-time virtual horse races. Its first public appearance was at the IVA07 conference where it participated in the GALA07 contest, having to report on a race script supplied on the spot. The successful result, being a jury award shared with another contestant as well as the public award, combined with the modular design inspired further development into a platform for transforming domain-specific information into a compelling live report by an embodied agent.

2 System description

The following is a brief description of the system. A more elaborate version is available in [1].

The first step in the process is the transformation of the domain-specific information into meaningful events. Different domains ask for a different method of determining what is meaningful and consequently, for every domain, a domain-specific Input Analysis Module (IAM) exists. These modules are required to deliver the events in a specific format, enabling further processing.

For all the events, the Mental Model Module (MMM) determines the emotional impact of each event based on the event information and personality parameters of the reporter. It also maintains a general emotional state which is the 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), with the prioritization based on the importance and decay factor parameters the IAM has linked to the events. Updating the importance using the decay factor ensures that the events reported on first are indeed the

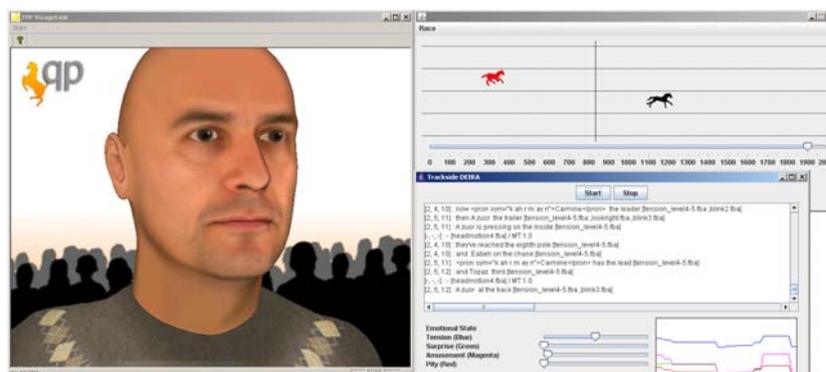


Figure 1: Screenshot of the system

most important ones. 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 retrieves from the EQ. To prevent repetition in verbal content, a history of utterances is maintained. The grammar provides a rich vocabulary to report on each event, and is easy to expand or adjust for other purposes.

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

The facial expression of the agent is handled by the Facial Animation Module (FAM) in two ways. Secondary head animations (like saccadic eye movements and small head motions) are triggered at fixed intervals. When reporting on a specific event the emotional state of the reporter is used to generate an appropriate primary animation like smiling and frowning.

When all aspects of the output have been determined, an Output Module (OM) feeds the text plus utterance characteristics and animation data to an external application responsible for actually displaying the animated model and cooperating with a Text-To-Speech (TTS) engine to generate lip-synced speech.

The external applications we currently support are the Haptek Player¹ as well as an application based on Visage², both capable of using a TTS engine such as Nuance's RealSpeak™ US English voice which we have used in most applications. Alternatively, audio-only output using Cloudgardens³ TalkingJava is also presently supported.

3 Modularity

At every step of development we have kept modularity, flexibility and interchangeability of different parts of the system in mind. The context-free grammar used by the TGM is contained in a human readable text-file and is completely language independent. Combined with being able to use more or less any TTS engine, audio output is truly language independent. The OM is furthermore easily replaceable to support different external applications, with all the different output modules being selectable *at runtime*.

The most recent work we have done to explore the abovementioned characteristics has been to add support for the RoboSoccer domain as well as support for the Virtual Storyteller project⁴ at the University of Twente. The adaptations necessary to support these domains were completed in 13 and 22 hours, respectively, which serve as support for our claims of modularity and adaptability.

The ability to adapt the system to diverse domains enables it to be used in a multitude of applications which can benefit from an embodied agent conveying situational information.

4 Demonstration

DEIRA is able to run demonstrations of the virtual storyteller, the robosoccer and the horse racing module, with demonstrations of the different instances lasting about 3 minutes per instance. This can be done using a sufficiently powerful laptop which is very preferably connected to external audio and video outputs.

References

- [1] F.L.A. Knoppel et al. Trackside DEIRA: A Dynamic Engaging Intelligent Reporter Agent. *In 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, pp. 112-119

¹ <http://www.haptek.com>

² <http://www.visagetechnologies.com>

³ <http://www.cloudgarden.com>

⁴ <http://hmi.ewi.utwente.nl/showcase/The%20Virtual%20Storyteller>