Conversing with Franco, FOCAL's Virtual Adviser

Michael Broughton, Oliver Carr, Paul Taplin, Dominique Estival, Steven Wark, Dale Lambert
Human Systems Integration group
Command & Control Division
DSTO
[firstname].[lastname]@dsto.defence.gov.au

Abstract

This paper presents the current dialogue capabilities of a Virtual Conversational Character (VCC) named Franco, which is an integral part of the Future Operation Centre and Analysis Laboratory (FOCAL). The VCC is used in a "Virtual Adviser" (VA) role to present multimedia information from pre-generated marked up scripts, and to engage in unscripted conversation with human users. The Virtual Adviser can run and control other applications within the virtual environment, allowing a "natural" interface between the user and complex information systems. The Virtual Adviser has been demonstrated with the Binni scenario used by the DARPA CoAX project.

1 Introduction

Presenting information by VCCs has been demonstrated by various systems in the open literature. Examples include "Ananova", a virtual newscaster that presents text based information via a speaking, digital talking head (Ananova, 2002). Another system is the Personalised Plan-based Presenter (PPP), a cartoon style, embodied character that presents information by emulating human-human communication techniques. The PPP combines pointing, head movements and facial expressions to draw the viewer's attention to the information being presented. (Andre, Rist and Muller, 1998).

One of the key new technologies we are developing for the Future Operations Centre Analysis Laboratory (FOCAL) is a "talking head" named Franco. In addition to the presentation of information, Franco is able to converse with users through the integration of automated speech recognition (ASR) and dialogue management systems. Franco is capable of conversing with the user on selected topics within a limited domain.

FOCAL is a prototype future command centre, based around a 150°, semi-immersive, collaborative, virtual reality environment. It is currently used for developing, integrating, and evaluating technologies potentially suitable for command centres of the future.

This work is part of ongoing research under the "Virtual Adviser" (VA) and "Virtual Interaction" (VI) work packages of the FOCAL task, currently underway in the Command and Control Division of the Defence Science and Technology Organisation (DSTO) at Edinburgh, South Australia.

2 Introducing… Franco

Franco, shown in Figure 1, is an animated, 3-dimensional "talking head" model (Taplin et al, 2001), developed using Alias|Wavefront™ Maya™, and making use of its real-time animation rendering engine and high-level scripting language. The Festival speech synthesis system (Festival, 2002) performs text-to-speech (TTS) conversion to provide Franco's voice. Franco is capable of delivering prepared information, such as a briefing or slide show, or interacting with users conversationally. Independent autonomous behaviour such as blinking and minor head and eye movements, provide increased realism. An SGI Onyx3400 with IR2 graphics engine is used as Franco's graphics and TTS engine, and renders him into an immersive VR environment.

Figure 1. VCC named Franco – FOCAL's first Virtual Adviser

Franco is controlled from a Java implemented console application, that can accept piped, manually entered, or automatically generated input. Information to be presented by Franco is entered as marked up text with embedded behaviours, system calls and links to multimedia files. Further detail relating to presented multimedia and embedded behaviours can be found in section 4.3. A schematic summarizing the implementation details is shown in Figure 2.

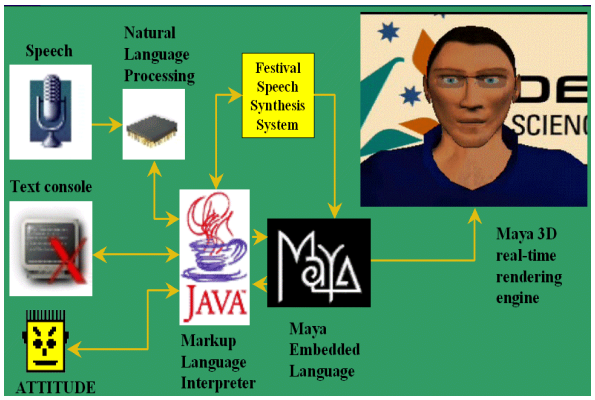

Figure 2: Schematic of Virtual Adviser

3 Knowledge Domain

Our current Virtual Adviser is designed to interact with the user within a limited domain relating to military aircraft, ships and geographic information in a demonstration scenario based around the CoAX Binni scenario (Allsopp et al, 2002), used to showcase FOCAL's capabilities. In this fictional scenario an Australian ship in a UN mandated peacekeeping role is attacked, and the role of the Virtual Adviser is to allow the user to gain a better situation awareness of what has occurred and to assist in the generation of a medical evacuation of the injured. The information has been sourced in the scenario from publicly available material, including the internet. The Adviser can also provide additional information about the FOCAL facility and has fictitious personal interests. For questions that go beyond the current knowledge base, an ELIZA style system has been implemented (Weizenbaum, 1966), where the question is cheekily mimicked back to the user.

4 Implementation of Conversation System

To create an initial conversational system using Franco, a laptop computer with Dragon NaturallySpeaking™ installed, has been utilised (Dragon, 2002). Dragon NaturallySpeaking is a speaker-dependent, automated speech recognition (ASR) system designed for PCs and has been chosen because of its good developer support, high recognition accuracy, and availability. Connected to the USB port is a high quality, head-worn, noise-cancelling microphone. Supporting software has been developed using Python (Lutz, 1996), which is an interpreted, interactive, object-orientated programming language. Python is portable across most computing platforms, and has an interface to Dragon NaturallySpeaking via NatLink (Gould, 2001). NatLink allows powerful macros to be developed in Python for Dragon NaturallySpeaking, including the ability to create grammars, which instruct the ASR what to listen for, and to generate callbacks when speech matching the grammar is recognised. The grammar can include optional and alternate words, which are particularly relevant for natural language processing. Two separate Python programs are used, one running on the PC and the other on the Onyx. Section 4.1 and 4.2 explains these two programs.

4.1 Background Noise Filtering

The program running on the PC utilises NatLink for integration with Dragon NaturallySpeaking and allows the specification of grammars. We have designed the grammar to eliminate background noises from entering the system. The grammars specified in this program allow all sentences to be parsed, but are very restrictive on single word utterances by the user. The only single words parsed by the grammar are the defined key words and other common single word responses such as 'Yes' and 'No'. All other single words are rejected, as they are most likely an insertion error caused by unwanted noise and the forced insertion mode of the ASR. With the current grammar implementation, if these unwanted singular word outputs are not eliminated, they typically result in a perceived utterance that cannot be handled by the current system.

Along with insertion errors, ASR systems also make substitution and deletion errors. Substitution errors occur when the correct word is incorrectly recognised and substituted with an incorrect word, while deletion errors occur when an utterance provides no response from the ASR system. Due to the key word technique we have adopted, substitution and deletion errors have not had a significant impact on performance, due to high recognition accuracy of the key words. This can be achieved by the user undertaking additional training of the ASR system, specifically targeting the key words.

4.2 Natural Language Processing

The program running on the Onyx handles dialogue management and response generation. The dialogue management is achieved through a series of strings in the form of regular expressions. The regular expressions are composed of key words, as opposed to complete sentences, to provide a natural language processing ability.

The key word technique was chosen for a number of reasons when compared to a complete grammar. First, it provides greater flexibility in the phrasing of questions by the user, therefore increasing robustness. Secondly, as we were using a very large active vocabulary, speaker dependent ASR system, the key word technique is more tolerant of erroneous words resulting from the ASR system. This is particularly relevant, as the ASR system was not tightly coupled to a grammar. Tight coupling between the ASR system and the grammar typically produces increased recognition accuracy at the expense of question flexibility.

When the relevant key words are correctly recognised, suitable responses can be generated from data pertinent to our demonstration scenario. The regular expressions are programmed in a hierarchical manner to support context sensitivity. Users can ask a general question about a particular military platform, then progressively ask more detailed questions as information is presented. An example is shown below, where (1) is a general question asked about the aircraft, and (2) and (3) has the user requesting more specific information.

(1) What do we know about the Sukhoi Su-27?

The Sukhoi Su-27 is a soviet-designed multi-role fighter designed as …

(2) What is its takeoff distance?

The Sukhoi Su-27 can takeoff in …

(3) And its flying range?

The Sukhoi Su-27 has a flying range of …

This principle can be extended to multiple levels of nesting, enabling further querying of returned information. When utterances cannot be matched within the regular expressions of the current context, the parent context becomes active again. The user can also shortcut the hierarchy and simply ask a question by using an attribute in combination with the asset name. An example is demonstrated in (4) below.

(4) What is the takeoff distance of the Sukhoi Su-27?

The previous examples have illustrated questions to retrieve single attributes of an aircraft. Aircraft specifications can also be displayed in a tabular format such that relevant information is provided in a single window. In this instance Franco does not read out the information in the table.

The grammar supports alternate words, such as abbreviated versions or nicknames of military platforms in our scenario.

4.3 Multimedia Output

In addition to spoken output from the Virtual Adviser, we also present imagery and multimedia to enhance the presented information. For example, we display images of particular military assets to supplement the textual information, provide location details on digital maps, and provide animations or movie clips where appropriate, to increase the richness of information presented to the user. The Virtual Adviser can also be used to launch applications within the computing environment in response to user or information cues. This allows the Virtual Adviser to assume the role of a virtual assistant for the user, and integrate painlessly with other technologies in FOCAL.

When other visual media is presented in conjunction with spoken output, Franco glances or turns his head toward the newly launched application to direct the users attention to the additional material being presented. Where the spoken utterances contain several sentences of information relating to the visual media, Franco glances back at the visual information at the start of each new sentence maintaining engagement with the users.

4.4 Franco's Personal Information

In addition to the professional data stored in the knowledge base, Franco also stores personal data, including his name, birth date, personal interests and underlying architecture information. Questions not captured in the main regular expression set are passed to the 'unknown question' handler that provides a selection of randomly assigned phrases to help the user rephrase their question. Additional regular expressions have been developed in an ELIZA style, that help trap asked questions that fall outside the primary knowledge domain.

I think you need a shave

Why do you think I need a shave?

4.5 A VCC with ATTITUDE

One of the main thrusts of the research effort in FOCAL is to automate the mundane information gathering, management, and reasoning processes to allow the user to deal more effectively with the dynamic, uncertain, high risk, and time critical military environment. In FOCAL this task will be handled by ATTITUDE, a multi-agent architecture developed at DSTO, capable of representing and reasoning with uncertainty and about multiple alternative scenarios. In the long term it is proposed to also use ATTITUDE as the dialogue management engine for the Virtual Adviser, and first steps along this path have been taken by demonstrating control of Franco by ATTITUDE running an implementation of ELIZA.

5 Future Directions

The current implementation of our interaction with Franco is relatively simple, primarily used to demonstrate simple grammars and to explore interaction capabilities with a Virtual Adviser. The next version will incorporate more sophisticated language processing by using Regulus (Rayner, 2001), a bi-directional language-processing component. Regulus, also referred to as Open NLP, is available as open source software. Initially, we intend utilising Regulus for natural language processing before we become involved with the language generation capabilities. We will also be implementing and investigating Nuance® 8.0 to provide our ASR capabilities (Nuance, 2002). Nuance has several attractions: primarily it is a speaker-independent system; has recently incorporated Australian-New Zealand English acoustic language models, and Regulus compiles into Nuance grammars. We are also investigating new TTS technologies, including rVoice™ by Rhetorical Systems (rVoice, 2002) and Nuance TTS engines. rVoice has the attraction of providing an Australian female voice developed by Appen (Appen, 2002), while Nuance TTS provides a male UK synthesised voice.

Measuring the cognitive value and user acceptance of the talking head within the FOCAL environment constitutes part of the planned research for the development of this project.

Acknowledgements

The authors wish to thank the Chief of C2D, and the Director of Information Sciences Laboratory, for sponsoring and funding this work. They wish to acknowledge the work of Andrew Zschorn in implementing ELIZA in ATTITUDE, and to thank the other members of the HSI

group in C2D for their constant and invaluable help with the FOCAL project.

References

Allsopp, D., Beautement, P., Bradshaw, J., Durfee, E., Kirton, M., Knoblock, C., Suri, N., Tate, A. and Thompson, C. 2002. Coalition Agents Experiment: Multi-Agent Co-operation in an International Coalition Setting, *Proceedings of the Second International Conference on Knowledge Systems for Coalition Operations (KSCO-2002),* Toulouse, France, pg 23-24 April 2002.(See also:http://www.aiai.ed.ac.uk/project/coax/).

Andre, E., Rist, T. and Muller, J. 1998. Integrating Reactive and Scripted Behaviours in a Life-Like Presentation Agent, *Proceedings of the Second International Conference on Autonomous Agents*, pg 261-268.

Ananova. 2002. http://www.ananova.com, accessed August 2002.

Appen. 2002. http://www.appen.com.au, accessed August 2002.

Dragon. 2002. Dragon NaturallySpeaking Voice Recognition, http://www.dragontalk.com/NATURAL.htm, accessed August 2002.

Festival. 2002. The Festival Speech Synthesis System, The Centre for Speech Technology Research, University of Edinburgh, http://www.cstr.ed.ac.uk/projects/festival/, accessed August 2002.

Gould, J. 2001. Implementation and Acceptance of NatLink, a Python-Based Macro System for Dragon NaturallySpeaking, *The Ninth International Python Conference*, March 5-8, California.

Lutz, M. 1996. Programming Python, O'Reilly & Associates, Inc.

Nuance. 2002. http://www.nuance.com/ accessed August 2002.

Rayner, M., Dowding, J. & Hockey, B. 2001. A Baseline Method for Compiling Typed Unification Grammars into Context Free Language Models. *Proceedings of Eurospeech 2001*, pg 729-732, Aalborg, Denmark.

rVoice. 2002. Rhetorical Systems, http://www.rhetoricalsystems.com/rvoice.html, accessed August 2002.

Taplin, P., Fox, G., Coleman, M., Wark. S. and Lambert, D. 2001. *Situation Awareness Using a Virtual Adviser*, Talking Head Workshop, OzCHI 2001, Fremantle, WA.

Weizenbaum, J. 1966. Eliza – A computer Program for the Study of Natural Language Communication between Man and Machine, *Communications of the ACM*, Volume 9, Number 1, pg 36-45.