Spoken Dialogue for Virtual Advisers in a semi-immersive Command and Control environment

Dominique Estival, Michael Broughton, Andrew Zschorn, Elizabeth Pronger
Human Systems Integration Group, Command and Control Division
Defence Science and Technology Organisation
PO Box 1500, Edinburgh SA 5111
AUSTRALIA
{Dominique.Estival, Michael.Broughton, Andrew Zschorn}@dsto.defence.gov.au

Abstract

We present the spoken dialogue system designed and implemented for Virtual Advisers in the FOCAL environment. Its architecture is based on: Dialogue Agents using propositional attitudes, a Natural Language Understanding component using typed unification grammar, and a commercial speaker-independent speech recognition system. The current application aims to facilitate the multi-media presentation of military planning information in a semi-immersive environment.

1 Introduction

In this paper, we present the spoken dialogue system implemented for communicating with the virtual advisers (VAs) in the Future Operations Centre Analysis Laboratory (FOCAL) at the Australian Defence Science and Technology Organisation (DSTO). We are experimenting with the use of spoken dialogue with virtual conversational characters to access multi-media information during the conduct of military operations and in particular to facilitate the planning of such operations.

Unlike telephone-based dialogue systems (Estival, 2002), which are mainly created for new commercial applications, dialogue systems for Command and Control applications (Moore et al. 1997) generally seek to simulate the military domain and therefore require an understanding of that domain.

2 Using Virtual Advisers in FOCAL

FOCAL was established to "pioneer a paradigm shift in command environments through a superior use of capability and greater situation awareness". The facility was designed to experiment with innovative technologies to support this goal, and it has now been running for two years.

FOCAL contains a large-screen, semi-immersive virtual reality environment as its primary display, allowing vast quantities of information to be displayed. Our current VAs can be described as 3-dimensional "Talking Heads", i.e. only the head and upper portions of the body are represented. They can display expression, lip-synchronisation and head movement, along with certain autonomous behaviours such as blinking and gaze (Taplin et al., 2001). These factors all combine to add life-likeness to the VAs and create more engaging interaction with users.

Presenting information via a Talking Head has been commercially demonstrated by the virtual newscaster “Ananova” (Ananova, 2002). Embodied characters are also being developed and include the PPP (Andre, Rist and Muller, 1998) and Rea (Cassell, 2000). PPP is a cartoon style Personalized Plan-based Presenter that combines pointing, head movements and facial expressions to draw the viewer’s attention to the information being presented. Rea is a virtual real-estate agent that takes an active role in conversation, she nods her head to indicate understanding of spoken input, or can raise her hand to indicate a desire to speak.

Several VAs have been implemented for FOCAL, each having a particular role or knowledge expertise. For example, one adviser may have specialist knowledge relating to legal issues, another may have information relating to the geography of a region. Each VA has a different facial appearance, voice and mannerisms.

To demonstrate and evaluate the performance of VAs (and of the other FOCAL projects), a fictitious scenario has been developed that incorporates key elements of military planning at the operational level (see section 8). The VAs provide information rich briefs through the combined use of spoken output via Text-to-Speech (TTS) and multimedia. Relevant questions can be asked at the end of the briefs through the use of spoken dialogue.

3 Previous implementation: Franco

As described in (Taplin et al., 2001) the first VA in FOCAL, named Franco, was also an animated 3-dimensional "Talking Head" model, intended to either deliver prepared information, such as a briefing or slide show, or to interact conversationally with users. To demonstrate the conversational functionality (Broughton et al., 2002), it was implemented with a commercial speaker-dependent automated speech recogniser (ASR), Dragon NaturallySpeaking™. The Natural Language understanding component was implemented in NatLink (Gould, 2001) and a simple user-driven dialogue management, based on key-word recognition and nesting of dialogue states to provide context, was also implemented in Python.

Franco has been successful in demonstrating the proof-of-concept of a VA in the FOCAL environment. Answering spoken questions about specific military assets and platforms, it also permits the display of other types of information such as pictures, animated video clips, tabular information from a database, and location details on digital maps.

4 Improvements

Although Franco was successful in demonstrating the potential usefulness of a VA in a Command and Control environment for operational planning, it suffers from certain limitations which we are now addressing in a follow-up project.

The first limitation, and the easiest to remedy, was the unnaturalness of the synthetic voice we had given Franco. For greater effectiveness, we had to provide our VA with a more natural voice and with an Australian accent. We chose the new Australian TTS voice from Rhetorical, developed by Appen (rVoice, 2002). This required making some changes, some of them relatively important, to the interface with the talking head model to achieve lip-synchronisation, but that aspect of the work will not be addressed in this paper.

The second limitation was the relative rigidity of the dialogue management strategy we were using. The alternative approach we have developed is to create Dialogue Agents implemented in ATTITUDE. This is described in section 6.

The third limitation was due to the speaker-dependent nature of the ASR. While a speaker-dependent ASR allows greater flexibility in the input to which the VA can respond, we wanted to develop a system which could not only be demonstrated by the few people who have trained the speech recogniser, but where visitors themselves could be participants and could interact with the VA. Switching to a speaker-independent ASR led us to radically modify our Spoken Language Understanding component, and this is described in section 7.

The new implementation we describe here has allowed us not only to address those three limitations, but also to alter fundamentally the architecture of the system, opening up the dialogue management components to control and interaction by other tools and agents in the FOCAL environment. The resulting system is now fully modular and provides scalability as well as flexibility.

This new implementation allows us to focus our research into dialogue management issue, to investigate the use of ATTITUDE for dialogue management and to experiment with more natural language input.

5 Integration

Communication between the various components of the system (speech recogniser, dialogue control, virtual adviser control and multimedia display) is now achieved with the CoABS (Control of Agent Based Systems) Grid infrastructure (Global InfoTek, 2002). The CoABS Grid was designed to allow a large number of heterogeneous procedural, object-oriented and agent-based systems to communicate. Using the CoABS Grid as our infrastructure has allowed us to integrate all the components of the dialogue system and it will provide an easy way to integrate other agents and a variety of input and output devices. Communication between CoABS agents is accomplished via string messages.

6 Dialogue Management with ATTITUDE

ATTITUDE is a multi-agent architecture developed at DSTO, capable of representing and reasoning both with uncertainty and about multiple alternative scenarios (Lambert, 1999). It is a multi-agent extension of the MetaCon reactive planner developed for control of phased array radars on the Swedish Airborne Early Warning aircraft (Lambert and Relbe, 1998). ATTITUDE has some similarities with Prolog and other logic programming languages as well as with AI research on blackboard and multi-agent architectures. Because ATTITUDE was designed specifically to support the programming of reactive systems, it possesses powerful facilities for handling interactions of the internal system entities, both with each other and with the external world.

ATTITUDE is very high-level, weakly-typed, and thanks to the agent paradigm, it produces loosely coupled and modularised systems. For these reasons, and because ATTITUDE implements reasoning about *propositional attitudes*, it provides a very attractive framework in which to develop and express dialogue management control strategies. It is worth emphasizing here that ATTITUDE is not merely a notation to represent speech acts or communicative acts between agents, but that it is actually the programming language and environment in which both the agents themselves and the control structure for interaction between the agents are implemented and executed.

Because ATTITUDE has never been used for this purpose before, this is an interesting area of research in itself, and one of the goals of the project has been to see how ATTITUDE needs to be extended to implement dialogue management. Further, this allows us to investigate how far *attitude programming* (see section 6.2) can go towards expressing speech acts and communicative act type. However, we do not claim to employ the full power of propositional attitudes in our implementation yet. This is another area of research which we are now exploring. Neither are we yet at the stage where we could perform automatic detection of utterance type (Wright, 1998) or of dialogue act (Carberry and Lambert, 1999; Prasad and Walker, 2002).

6.1 Propositional attitudes

The ATTITUDE programming environment is so named because it utilises *propositional attitude instructions* as programming instructions (this has been dubbed *attitude programming*). Propositional attitudes are alleged mental states characterised by propositional attitude expressions, which are the means by which individuals relate their own mental behaviour to others'.
Propositional attitude instructions are of the form shown in (1).

(1) [subject][attitude][propositional expression]

In (1):

- [subject] denotes the individual whose mental state is being characterised;
- [propositional expression] describes some propositional claim about the world; and
- [attitude] expresses the subject's dispositional attitude toward that claim about the world.

6.2 ATTITUDE programming

When software agent *Mary* encounters the propositional attitude instruction "*Fred* **desire** [the door is closed]", *Mary* will issue a message to software agent *Fred* instructing *Fred* to desire that the door be closed. Similarly, when encountering the propositional attitude instruction "*I* **believe** [the sky is blue]", *Mary* herself will attempt to believe that the sky is blue.

An important characteristic of ATTITUDE programming is that each propositional attitude instruction either succeeds or fails, possibly with side effects, depending upon whether the recipient agent is able to satisfy the instructional request. As each propositional attitude instruction either succeeds or fails, the execution path selected through a network of propositional attitude instructions (routine) is determined by the successes and failures of the propositional attitude instructions attempted along the way. The control structure is therefore governed by a *semantics of success.*

Computational routines for a software agent arise by linking together particular choices of propositional attitude instructions. These networks of propositional attitude instructions then prescribe recipes defining the possible mental behaviour of a software agent.

6.3 The ATTITUDE Dialogue Agents

We have implemented a number of ATTITUDE Dialogue Agents. The main agent in our Dialogue Management architecture (shown in Figure 1) is the *Conductor*. It is the agent responsible for the flow of information between the other agents and it manages multi-modal interactions. The other agents, also described further in this section, are the *Speaker*, the *NLG* (Natural Language Generator), the *MMP* (Multimedia Presenter) and several *IS* (Information Source) agents. In addition to these agents, each dialogue state (see section 8) is also implemented as an ATTITUDE agent, with its own set of routines.

As explained in section 6.2, each ATTITUDE agent's behaviour is programmed as a set of routines

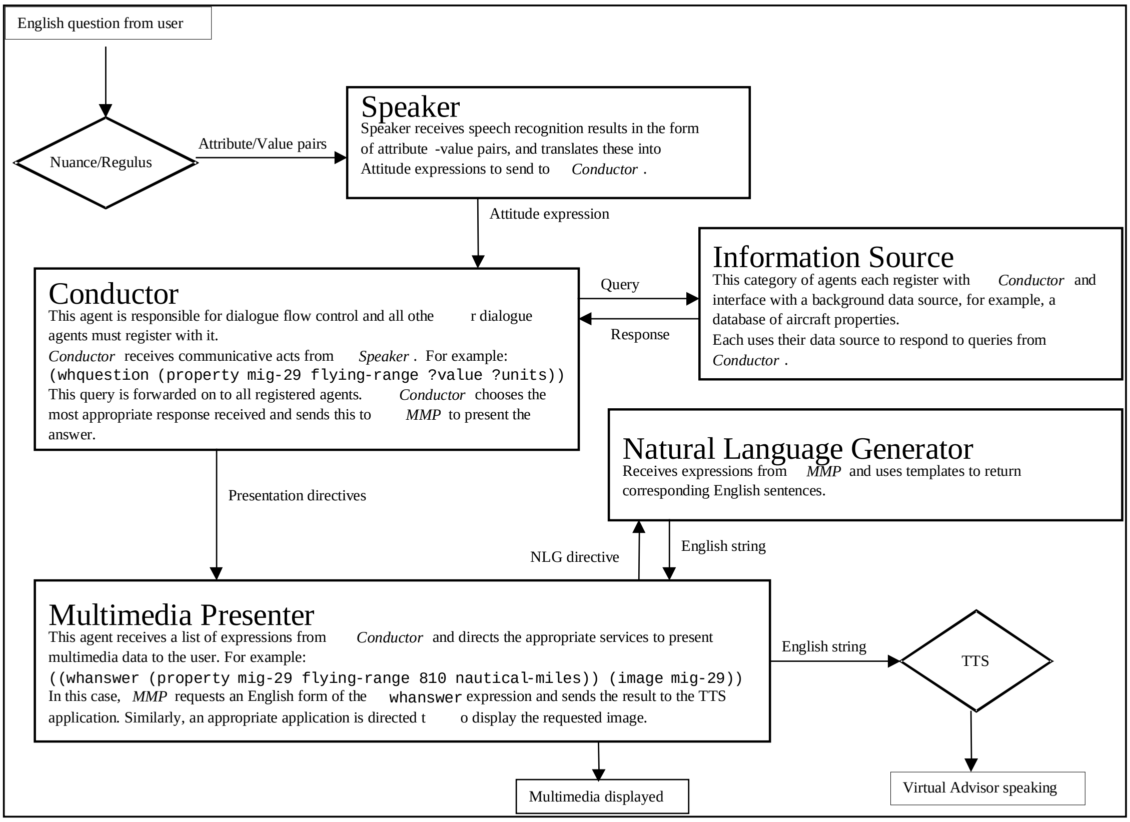
The interaction between the ATTITUDE Dialogue agents is shown in Figure 1, in which the frame around the ATTITUDE agents can be interpreted as representing the CoABS grid.

Speaker Agent

When speech from the user has been detected and recognised, the attribute-value pairs for that utterance (see section 7) are sent to *Speaker*. *Speaker* takes that information and produces a corresponding ATTITUDE expression, which is then forwarded to *Conductor*.

The linguistic coverage of the system is determined by the grammars which are available at each dialogue state. For now, the coverage is limited to a set of utterances appropriate for the briefing scenario described in section 8. These were used to define the Regulus1 grammars from which the Nuance grammars are compiled. We are now planning to move from Regulus1 to Regulus2, which will allow us to derive dialogue state grammars from a large English grammar using the EBL strategy described in (Rayner et al., 2002b)

Figure 1. Dialogue with ATTITUDE



Conductor Agent

Conductor takes an ATTITUDE expression from *Speaker* and forwards it on to all the *IS* agents that have registered with it. It then waits for all the responses to come back from those agents, in the form of lists of expressions.

Every response *Conductor* receives is put into its knowledge base, along with some extra information:

- Sender: which IS agent sent the response.
- In-Reply-To: which previous communicative act this is a response to.
- Strength: whether every expression of the response is 'strong' (the sender believes it is either absolute truth or absolute negation) or if one or more is 'weak' (the sender believes it is neither absolute truth nor absolute negation).
- Bound-State: if there are any free variables in the response, or if it is fully ground.
- Unifiability: whether one or more of the expressions in the response is of the same form as *Speaker's* initial expression.

The final expression in *Conductor*'s knowledge base is as shown in (2).

(2) (response ?in-reply-to ?sender ?strength
  ?bound_state ?unifiability ?content)

Given the initial expression from *Speaker* and the replies it receives from the IS agents, *Conductor* chooses the 'best' response. For example, a response that is strong, fully ground and unifies with *Speaker's* expression is deemed to be more relevant and informative than a response that is weak and contains free variables. *Conductor* forwards this response to *MMP*.

Multimedia Presenter (MMP)

MMP iterates through the list of expressions sent by *Conductor* and presents each expression to the user. *MMP* recognises classes of expressions and chooses to present them using certain media. For example, some expressions are instructions to change the VA head model, while others are to be translated into English sentences and spoken by the VA. For the latter function *MMP* uses *NLG* (see below).

Other media through which *MMP* can choose to present the information contained in the expressions include: imagery from a database (e.g. pictures of military platforms, or of strategic locations), video clips, images from weather or radar information sources, virtual video, 3-dimensional virtual battle space maps, textual information and audio.

Natural Language Generator (NLG)

For now, *NLG* uses templates to transform ATTITUDE expressions into English. For example, the instruction in (3) provides two possible responses for the ATTITUDE expression specified:[1]

```
(3) (property ?asset overview ?value text)
    whanswer priority 10
   ( ( response 1 ("The "?asset" is a "?value".") )
   ( ( response 2 ("I understand that the "?asset" is a
      "?value".") )))
```

When *NLG* is first requested to generate the English output for the expression in (4.a), intended to be a communicative act of type *whanswer*, it uses the template given in (4.b), corresponding to "response 1" in (3), to produce the English answer given in (4.c).

(4.a) (property mig-29 overview "Russian multi-role fighter" text)
b. ("The "?asset" is a "?value".")
c. The Mig-29 is a Russian multi-role fighter.

When *NLG* is requested a second time to generate the output for (3), it uses the template in (5.a), corresponding to "response 2" in (3), to produce the English answer given in (5.b).

(5.a) ("I understand that the "?asset" is a "?value".")
b. I understand that the Mig-29 is a Russian multi-role fighter.

Thus *NLG* cycles through the list of templates for appropriate responses. Priorities can also be given to templates, enabling *NLG* to use general templates together with more specific and tailored ones.

It is clear that template-based language generation is too rigid for fully natural dialogues, and we intend to explore more flexible techniques after we implement a wider coverage English grammar; however, it has so far been sufficient for

[1]Variables are denoted with "?", while text strings (to be sent to speech synthesis, or displayed on a slide) are between double quotes, "".

our purposes, namely to demonstrate and investigate agent-based dialogue management.

Information Source Agent (IS)

The *IS* agents, e.g. a Weather Agent or a Platform Capabilities Agent, can answer users' questions, either by using their own internal knowledge base or by accessing external Information Sources, such as a weather information server, or a database of military assets. All *IS* agents register with *Conductor*, and when an expression is sent by *Speaker*, all *IS* agents try to respond to it.

By using the CoABS Grid as the infrastructure and implementing the agent with ATTITUDE, we leave the architecture extremely flexible and scalable (Kahn and Della Torre Cicalese, 2001). For instance, it is possible to increase the amount of information at the system's disposal during run-time by launching a new *IS* agent and by adding some templates to *NLG*.

6.4 Dialogue design

For now, the dialogue is specified as a finite state machine and is still very much system directed. In the briefing application (see section 8.1), the VAs first "push" the information that needs to be presented, as briefing officers do in a normal briefing. Some of the information is also presented using visual aids, such as power point slides and maps for specifying location information. The information to be presented and the media to be used are determined by the agent for that particular dialogue state.

The VA then allows users to ask questions to repeat or clarify particular points, or to gain additional information.

7 Spoken Language Processing

7.1 Speaker-independent speech recognition

As stated in section 4, one of the main motivations for moving from a speaker-dependent to a speaker-independent ASR was to allow visitors in FOCAL the possibility of using the system themselves, rather than relying on a small set of trained individuals to run demonstrations. We chose to use the Nuance Toolkit (Nuance, 2002) for several reasons: besides its reliability as a speaker-independent ASR for both telephone and microphone speech, Nuance 8.0 provides Australian-New Zealand English, as well as US and UK English, acoustic language models. Even more importantly for our purposes, Nuance grammars can be compiled from Regulus, a higher-level language processing component which has already been used to develop several spoken dialogue systems in different domains (Rayner et al., 2001, Rayner and Bouillon, 2002).

7.2 Spoken Language Understanding

Following our decision to move from a speaker-dependent to a speaker-independent ASR, we decided to use Regulus to implement our Natural Language Understanding component. Regulus is an Open Source environment which compiles typed unification grammars into context-free grammar language models compatible with the Nuance Toolkit. It is "written in a Prolog-based feature-value notation and compiles into Nuance GSL grammars." (Rayner et al., 2002a). Regulus is also described in detail in (Rayner et al., 2001).

The main motivation for using Regulus is the usual one of greater efficiency due to the more compact nature of a unification grammar representation compared with a context-free grammar. In addition, using Regulus to define a higher level grammar, we are able to obtain as our semantic representation a list of attribute-value pairs, and this permits a more sophisticated processing of the information by the other agents.

Regulus also allows the development of bi-directional grammars, and we intend to make use of this functionality in later implementations of the *NLG* agent. However, for now, the grammars we have developed have been limited to recognition and understanding.

8 Current application implementation

8.1 Dialogue scenario

The scenario for the current application was developed by members of the Human Systems Integration (HSI) group and is grounded on their experience with, and observations of, military operational planning. It is based on a fictitious scenario developed for training (the examples given here have all been modified) and exemplifies the Joint Military Appreciation Process (JMAP) for military planning across the three services (Army, Navy and Air Force). A sub-scenario was chosen for the development of the spoken dialogue with the VAs.[2]

8.2 Dialogue flow

The structured nature of a military planning task such as this one makes it very easy to partition it into different stages, which can then be mapped to different dialogue states. In our dialogue script, each top-level dialogue state corresponds to a section of the planning exercise, given in (6).

(6) Commander's Initial Guidance
- CDF (Chief of Defence Forces) Intent
- Planning Guidance
- Constraints
- Restrictions
- Legal Issues
- Command and Control

These 6 top level dialogue states are then followed by an Overall Question Time.

The mixed-initiative nature of the system can be modelled in a finite state diagram, allowing for a) briefing-like system 'pushes', b) confirmation queries from the system and c) questions from the user. However, because the system is primarily agent-based, the dialogue can also evolve dynamically. For instance, once the system is in a 'question' state, the dialogue flow then allows users to ask a number of questions, until they are satisfied, and the dialogue can move to a different state.

Each of the top level dialogue states also corresponds to an *IS* agent with its own set of ATTITUDE routines. These agents register with *Conductor* and act as experts in their particular fields (e.g., the Legal Issues adviser). The agents contain knowledge which they use to answer questions posed to them by *Conductor*. All agents have the ability to keep track of which state (or topic) they are in. This allows not only *Conductor*, but also the other dialogue agents, to distinguish between providing the user with new information or information that has already been presented.

[2] This is the Commander's initial guidance to the Theatre Planning Group (TPG), which is part of the Mission Analysis section of JMAP.

8.3 Knowledge Representation

The current ontology developed for this application is only a small part of the larger Knowledge Representation ontology to be used throughout the whole FOCAL system. For now, we only represent the concepts needed in our small domain, and their relationships are translated into ATTITUDE statements, allowing agents to draw inferences. For example, if a user can ask the question given in (7.a), it will be translated into the list of attribute value pairs given in (7.b) and sent to *Speaker*. *Speaker* then translates these attribute value pairs into the ATTITUDE expression in (7.c) and forwards it on to *Conductor*.

(7.a) What department oversees negotiations with unions and industry?
b. [question whatquestion, concept negotiation, attribute oversee, obj1 department]
c. conductor desire (comm_act (negotiation oversee ?department) from speaker type whatquestion in-response-to null)

As described in section 6, when *Conductor* poses the question to the appropriate agents, they respond with the information in their knowledge base or information they can extract from a database. Agents store knowledge as **believe** statements such as the one shown in (8):

(8) I believe (negotiation oversee "department of workplace relations")

These **believe** statements are then unified with the propositions translated by *Speaker*, and if unification is successful, a reply is sent back to *Conductor*. Finally, *Conductor* passes the answer on to *NLG* to match a template and produce an English answer, for instance (9).

(9) The Department of Workplace Relations oversees negotiations with unions and industry.

An agent which has access to a database can also translate a user's question into the relevant database query to obtain the answer. An important issue under research concerns the automatic derivation of ATTITUDE statements from a pre-existing database.

8.4 Several different VAs

As explained above, each stage of the planning process is presented to the user by a particular VA with its associated *IS* agent and the VA then allows users to ask further questions. Besides their specialised knowledge, the VAs are differentiated through different head models, different TTS voices (male or female, different regional accents) and different personalities.

Once a dialogue state is completed and the user has no further questions, the VA for that state sends a message to *Conductor* to move to the next state. *Conductor* can then initiate the change in recognition grammar, voice for the next VA and model for the next VA head.

Having several VAs coming on at different stages to present different information allows for a VA to be specialised in a particular domain, just as real briefing officers are during a real military planning exercise.

For now, we only display one VA at a time, but we intend to experiment with having multiple VAs at the same time. The final state of the dialogue flow allows users to ask questions about any aspect of the planning process, and questions can be posed to all the VAs, so it would be natural for the users to see all the VAs at that stage.

8.5 Rapid Prototyping and Evaluation

The key word version developed previously (see Broughton et al., 2002) has been maintained as a rapid prototyping environment for evaluating new scripts and dialogues. It allows new dialogues to be quickly tested by entering suitable key words, sufficient to discriminate one question from another. This system proves faster for testing than the more precise method of grammar building. Multiple response strings can be generated, providing more naturalness for those interacting with the VAs on a regular basis. By rapidly prototyping questions and responses, we can test the intuitiveness of expected questions and the smoothness and timeliness of responses, particularly when presented combined with multimedia.

The implemented system described here has so far only been tested with other members of the group, but demonstrations to visitors and potential users will provide a more rigorous form of evaluation on an on-going basis. An evaluation phase for the project is scheduled for 2003-2004, during which time we will have access to more users and will be able to conduct more structured experiments.

9 Natural Interaction with VAs

In addition to the ASR and TTS systems previously discussed, other technologies can be combined into the overall system to increase naturalness of interaction, and we are investigating speaker recognition as well as a range of pointing technologies.

The need for a speaker recognition system has emerged with the move to a speaker independent ASR. With a speaker dependent ASR, users would load their individual profile before use, thus enabling the system to know who was using it. With a speaker-independent ASR, a speaker recognition system would allow the VAs to recognise who is talking to them and enable them to address known users by name. We plan to integrate within FOCAL the speaker recognition system which has been developed at DSTO (Roberts, 1998). This system uses statistical modelling techniques and is capable of both speaker identification (recognising users from a database of stored speech profiles) and speaker verification (verifying the identity of a particular user).

We are also proposing to use pointing techniques in combination with the speech and language technologies to build a multimodal system. Multimodal systems were originally demonstrated by Bolts (1980) and research is continuing across varied applications (e.g., Oviatt et al., 2000 and Gibbon et al., 2000). However, unlike systems such as MATCH (Johnston et al., 2002), where the issue is allowing multimodal interaction on portable devices with very small screens, in FOCAL we are concerned with ensuring that users get the full benefit of the very large screen and with allowing several users to interact at a distance from the screen. It is also worth mentioning that, unlike the interactive system described in (Rickel et al., 2002), which is concerned with training in a military environment, we are not trying to simulate a complete virtual world with embodied agents.

However, we propose to include traditional pointing technologies, such as the standard desktop mouse, through to 3-dimensional tracking systems for gaze, gesture and user tracking. This will involve integrating more complex language understanding, as information will need to be derived from both the user's utterance and from what is being pointed to. For example, to interpret an utterance such as (10) uttered while the user points to a location on a map, we need to perform reference resolution on "this region", and match that referent to the item being pointed at.

(10) What do we know about this region?

10 Conclusion

We have now implemented in FOCAL the infrastructure needed to perform spoken and multimodal dialogue with several VAs. This is of interest in itself, as it will allow us to continue our research on spoken language understanding and spoken dialogue systems and also to address issues of language generation which have for now been left aside. Already we have been able to move from a rigid dialogue control structure, with very constrained input, to a more flexible and scalable control structure allowing real connectivity between agents.

Having moved to a speaker-independent ASR, and taking advantage of the open source nature of Regulus, we intend to pursue research issues regarding robust processing of spoken input, such as using grammar specialisation from a corpus and devising techniques for ignoring parts of the input.

We have implemented a dialogue management architecture based on ATTITUDE agents which communicate with each other using propositional attitude expressions. Other agents can now be developed to perform additional functions, in particular to launch the display of other types of information and to interpret other types of input.

This will allow us to explore how spoken dialogue with VAs can be combined with other virtual interaction technologies (e.g., gesture, pointing, gaze tracking). In this respect, the next step in our project is the development of a full fledge *MMP* agent based on the framework described in (Colineau and Paris, 2003).

However, the work we have reported here must also be seen as part of the larger research programme undertaken within FOCAL. From this perspective, this work is of interest because it allows other members of the HSI group to pursue research in the usability of new technologies to perform the paradigm shift in command environments. In particular, this project is providing the support for further research into whether this way of presenting information is helpful in an operational command environment. It allows us to devise experiments to explore the crucial issue of trust in the information being presented, and how the way the information being presented can affect that trust.

Integrating spoken dialogue with planning tools will also allow us to explore whether VAs can help in military operation planning, and how best to use these tools.

Acknowledgements

We wish to thank the Chief of C2D, and the Director of Information Sciences Laboratory, for sponsoring and funding this work. We wish to acknowledge the work of Paul Taplin in integrating speech synthesis and lip-synchronisation, and the work of Benjamin Fry from the University of South Australia in developing the Regulus/Nuance grammars. Finally we wish to thank the other members of the HSI group in C2D for their constant and invaluable help with the FOCAL project.

References

Ananova. 2002. http://www.ananova.com.

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

Appen. 2002. http://www.appen.com.au.

R. A. Bolt. 1980. "Put-that-there": voice and gesture at the graphics interface. *Proceedings of the SIGGRAPH*, July, 262-270.

Michael Broughton, Oliver Carr, Dominique Estival, Paul Taplin, Steven Wark, Dale Lambert. 2002. "Conversing with Franco, FOCAL's Virtual Adviser". *Conversation Characters Workshop, Human Factors 2002*, Melbourne, Australia.

Sandra Carberry and Lynn Lambert. 1999. "A Process Model for Recognizing Communicative Acts and Modeling Negotiation Subdialogues". *Computational Linguistics*. 25,1, pp. 1-53

Justine Cassell. 2000. Embodied Conversational Interface Agents, *Communications of the ACM*, Vol. 43, No. 4, 70-78.

Nathalie Colineau and Cécile Paris. 2003. *Framework for the Design of Intelligent Multimedia Presentation Systems: An architecture proposal for FOCAL*. CMIS Technical Report 03/92, CSIRO, May 2003.

Dominique Estival. 2002. "The Syrinx Spoken Language System". *International Journal of Speech Technology*. vol.5. no.1. pp.85-96.

Michael Johnston, Srinivas Bangalore, Gunaranjan Vasireddy, Amanda Stent, Patrick Ehlen, Marilyn Walker, Steve Whittaker, Preetam Maloor. 2002. "MATCH: an Architecture for Multimodal Dialogue Systems". *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*. pp. 376-383. Philadelphia..

Dafydd Gibbon, Inge Mertins, Roger K. Moore (Eds.). 2000. *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. Kluwer Academic Publishers.

Global InfoTek Inc. 2002. *Control of Agent Based Systems*. http://coabs.globalinfotek.com.

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

Martha L. Kahn and Cynthia Della Torre Cicalese. 2001. "CoABS Grid Scalability Experiments". *Proceedings of the Second International Workshop on Infrastructure for Agents, MAS, and Scalable MAS*, Autonomous Agents 2001 Conference.

Dale A. Lambert and Mikael G. Relbe. 1998. "Reasoning with Tolerance". *2nd International Conference on Knowledge-Based Intelligent Electronic Systems*. IEEE. pp. 418-427.

Dale A. Lambert. 1999. "Advisers With ATTITUDE for Situation Awareness". *Proceedings of the 1999 Workshop on Defence Applications of Signal Processing*. pp.113-118, Edited A. Lindsey, B. Moran, J. Schroeder, M. Smith and L. White. LaSalle, Illinois.

Dale A. Lambert. 2003. "Automating Cognitive Routines", accepted for publication in the *6th International Conference on Information Fusion*.

R. Moore, J. Dowding, H. Bratt, J. Gawron, Y. Gorfu, A. Cheyer. 1997. "CommandTalk: A spoken-language interface for battlefield simulations". In *Proceedings of the Fifth Conference on Applied Natural Language Processing, pp 1-7*.

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

Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., Ferro, D. 2000. "Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions". *Human Computer Interaction*.

Rashmi Prasad and Marilyn Walker. 2002. "Training a Dialogue Act Tagger for Human-Human and Human-Computer Travel Dialogues". *Proceedings of 3rd SIGDIAL Workshop*. Philadelphia. pp.162-173.

Manny Rayner, John Dowding, Beth Ann Hockey. 2001. "A Baseline method for compiling typed unification grammars into context free language models". In *Proceedings of Eurospeech 2001*, pp 729-732. Aalborg, Denmark.

Manny Rayner, John Dowding, Beth Ann Hockey. 2002a. "Regulus Documentation".

Manny Rayner, Beth Ann Hockey, John Dowding. 2002b. "Grammar Specialisation meets Language Modelling". *ICSLP 2002*. Denver.

Manny Rayner and Pierrette Bouillon. 2002. "A Flexible Speech to Speech Phrasebook Translator". *Proceedings of the ACL-02 Speech-Speech Translation Workshop*, pp 69-76.

Jeff Rickel, Stacy Marsella, Jonathan Gratch, Randall Hill, David Traum, William Swartout. 2002. Toward a New Generation of Virtual Humans for Interactive Experiences. IEEE Intelligent Systems, 1094-7167, pp. 32-38.

William Roberts. 1998. "Automatic Speaker Recognition Using Statistical Models". *DSTO Research Report, DSTO-RR-0131*, DSTO Electronics and Surveillance Research Laboratory.

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

Paul Taplin, Geoffrey Fox, Michael Coleman, Steven Wark, Dale Lambert. 2001. "Situation Awareness Using a Virtual Adviser", *Talking Head Workshop, OzCHI 2001*, Fremantle, Australia.

Helen Wright. 1998. "Automatic utterance type detection using suprasegmental features". Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP'98). Sydney.