Speaking and acting – interacting language and action for an expressive character

Sandy Louchart[1], Daniela Romano[2], Ruth Aylett[1], Jonathan Pickering[1]

[1]Centre for Virtual Environments, University of Salford M5 4WT
[2] The Department of Computer Science, University of Sheffield, S1 4DP

Abstract: We discuss the FearNot! application demonstrator, currently being developed for the EU framework V project VICTEC. It details the language structure, content, interactions management and design of the FearNot! Demonstrator, as well as presenting the VICTEC project and its motivations. This paper also focuses on the different sets of Speech Act inspired language action lists developed for the project and discusses their use for an interactive language and action system for the elaboration of expressive characters.

1. Introduction

This paper discusses the language system being developed for the EU framework V project VICTEC[1] – Virtual ICT (Information and Communication Technologies) with Empathic Agents. This seeks to use virtual dramas created by interaction between intelligent virtual agents as a means of dealing with education for children aged 8-12 in which attitudes and feelings are as important as knowledge. The project thus focuses on Personal and Social Education, which includes topics such as education against drugs, sex education, social behaviour and citizenship. The topic specifically addressed by VICTEC is education against bullying.

An output of the project is the FearNot! [Figure 1] demonstrator, currently under construction. The overall interactional structure of this demonstrator alternates the enaction of virtual drama episodes in which victimisation may occur, and interaction between one of the characters in these dramas and the child user, who is asked to act as their 'invisible friend' and help them to deal with the problems observed in the dramatic episodes. The advice given by the child will modify the emotional state of the character and affect its behaviour in the next episode.

The FearNot! Demonstrator represents an intuitive interface between the virtual world and the child user. The characters appearing in the demonstrator have been modelled to be believable rather than realistic, with the use of exaggerated cartoon-like facial expressions. Evaluation to date [Wood et al 02] has shown that providing the narrative action is seen as believable, lack of naturalism is not perceived as a problem by prospective child users. FearNot! Draws upon feelings of immersion and suspension of disbelief, essential characteristics of Virtual Reality (VR) and Virtual

[1] This project is financially supported by the European union Framework V IST programme

Environments (VE), in order to build empathy between the child and the virtual character as the child explores different coping behaviours in bullying.

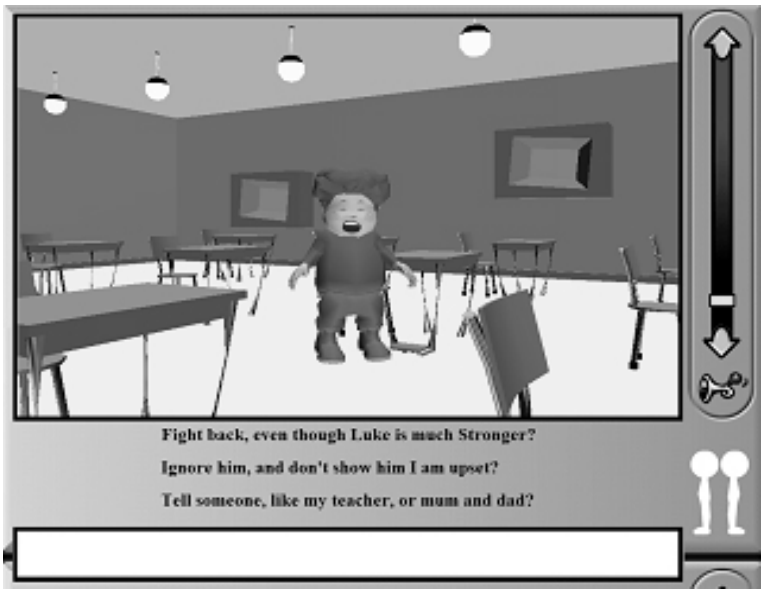


Figure 1: The FearNot! Demonstrator

2. Integrating language and action

Unlike most dialogue systems or *talking heads*, VICTEC mixes language interaction with physical actions. Bullying can be categorised as verbal, physical, or relational (manipulating social relationships to victimise), so that actions such as pushing, taking possessions and hitting must be modelled. Each character displayed in the FearNot! Demonstrator is provided with its own autonomous action selection mechanism, and the overall architecture is shown in Figure 2. An appraisal of events and the other characters is carried out, using the emotion-modelling system of Ortony, Clore and Collins [Ortony et al 88] and the resulting emotional state is combined with the character's goals and motivations to select an appropriate action. Thus a common representation for both physical actions and language actions is needed so that both can be equally operated upon by the action-selection mechanism.

This representation is provided by the concept of a *speech act* [Austin 62, Searle 69], defined as an action performed by means of language. Here, language is categorised by its illocutionary force, that is, the goal that the speaker is trying to achieve; the same view of action taken by an action-selection mechanism, and highly relevant to bullying scenarios. Speech Acts however work at a very high level of abstraction (e.g. assert, promise, threaten) and only a subset of those generally used are relevant to bullying scenarios. Moreover much of the subsequent work – such as that in Dialogue Acts [Bunt 81] – has taken place in language-only domains and does

not address the close relationship between speech and actions required for the VICTEC project. It was therefore decided to define a set of *language actions* in the spirit of speech acts, using a corpus of bullying scenarios constructed by school children using a story-boarding tool Kar2ouche [Kart2ouche].

Of course a speech act does not uniquely specify the utterance in which it is expressed – its locutionary form. Moreover it was created as an analytic tool, while the language system being created here must function in a generative capacity (see [Szilas 2003] for other work with this aim). In addition, language and other actions must form coherent sequences, accepted as such by the child users. The approach must also take account of cross-cultural language practices such as the specific language used in schools in the UK, Portugal and Germany, the countries of the project partners.

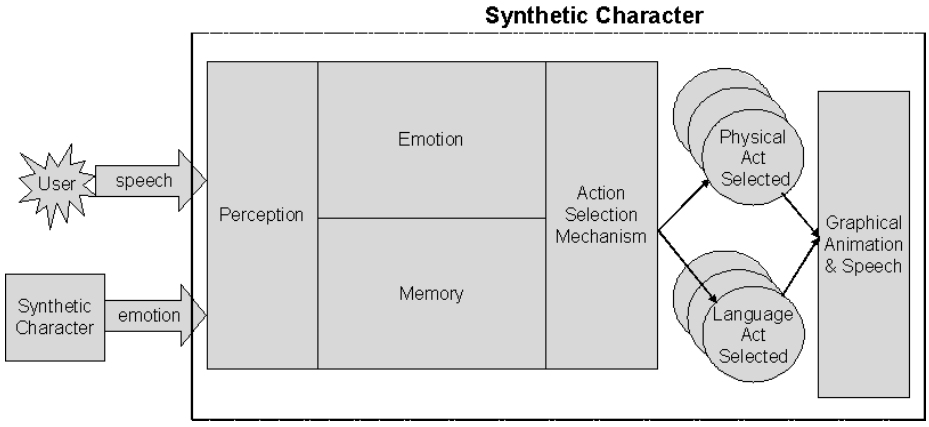


Figure 2: The Synthetic Character Architecture

Finally, there are two different contexts in which the language system must work. The first is *within* dramatic episodes in which characters interact with each other. The second is *between* episodes in which the character must interact with the child user.

2.1 From action to utterance

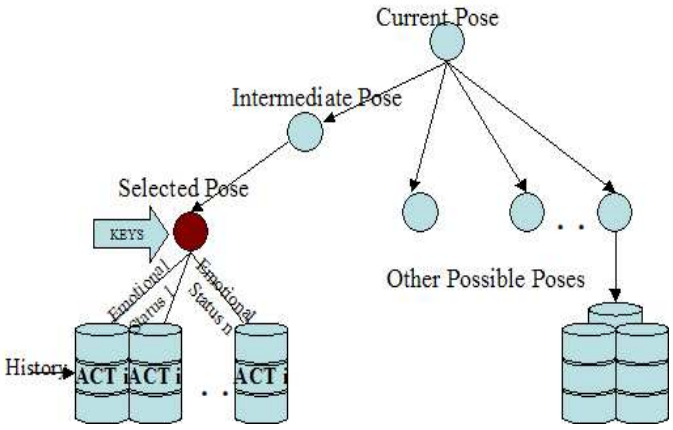
An action can be described as a collection of instances of: an **object** on which the action can be performed (those being a object of the environment or another characters), the **agent** performing the action, the action **priority** used to order and deal conflicting actions, the **context** in which the action is performed (i.e. location, props, internal goal, history of previous actions, topics), the **emotional status** of the character at that time, and the **utterance** (relating to the language action) that should be played, and the **animation** of the part of the body of the character involved and the gesture. The emotional status of the character will determines whether the action to be performed is implemented via language action, physical activity or both.

Assuming that the next action selected is physical, from a *current pose* of the character a series of animations are possible, but to reach the current select one it

might be necessary to introduce an intermediate pose that links the two (i.e. next action: walk to the door. Current pose: sitting. Intermediate pose necessary: stand up).

We can visualise this as a tree of behaviours where from a current state the next animation is possible only when the correct status of the character is reached and that action can began, requiring the introduction of an intermediate pose. See figure 3.

Figure 3: The Tree of Behaviours



In order to generate the utterance for a selected language action, it has been decided to use a shallow-processing approach, as originally used in ELIZA [Weizenbaum 66] and more recently in chat bots [Maudlin 94]. The rationale for this approach is that it takes little processing resource compared to a deep approach based on parsing and semantics, thus allowing the graphics engine the resource it needs to run in real-time. In addition, such systems can show surprising resilience in limited domains such as that of FearNot!, in which the language to be used is specific to the bullying scenarios. To prevent problems experienced with such systems in dealing with unexpected inputs, the FearNot! demonstrator will specifically drive the conversation in child-agent mode by using leading questions with a limited range of options for answer. Wizard of Oz studies are in progress to determine in more detail what language coverage will be required.

In agent-agent interaction, the language system starts with the language action generated by the action-selection system, which has the advantage of knowing exactly what action (language or otherwise) it is responding to. This indexes a group of utterance templates in which the previous utterance or physical action is used to fill in variable slots with an appropriate choice. For example, assume the utterance from the other agent was“ I like flowers”, the following group of utterance is selected: {I like ---- too, why do you like ---?, what do you find in ---?}. The first unused utterance here is: ”why do you like ---?” the dots are filled with the recognized object of the discourse in the user’s input: flowers. The generated character utterance is “why do you like flowers?”.

Child and character interaction is different. Here the action is not known, but must be inferred. The incoming text is matched against a set of language templates, and the language and action index is then taken as the starting point for the language action with which the agent must respond as discussed below. Since an objective is to retain control of this dialogue by keeping the conversational initiative with the character, the Finite State Machine structures discussed below can also be used to generate expectations about what language actions the child has produced.

3. The FearNot! Speech Act Knowledge-base

In order for the FearNot! Demonstrator to successfully meet VICTEC's evaluation objectives, it is crucial that continuity and coherence is maintained during interactions (contextualisation) between agents while insuring that the communication is engaged and led by an agent when agents and users interact together. This not only fundamentally affects the design of the language system, it also requires the design of two distinct sets of actions independent of each other as just discussed.

3.1 Action categorisation

A set of appropriate actions for bullying and victimization interactive scenarios has been identified. Those actions can be triggered and generate agent utterances according to their emotional states. As such a system is dealing with a number of actions and utterances, we have grouped the entire language content within three categories, **Help, Confrontation** and **Socializing** [Table1].

HELP
Ask for help / Offer help / Help question / Help advice / Help introduce to friend / Help talk to someone / Help invitation / Offer protection / Non assistance confirmation
CONFRONTATION
Order / Aggressive questioning / Do / Forbid / Defiance / Tease / accusations / Insult / Threat / Aggressive answer / Apology / Abandon action / Action / Hit / Lie / Steal / Obey / Deny / Ask why / Beg / Claim back / Leave / Struggle.
SOCIALISING
Greeting start / Topic introduction / Exclusion topic introduction / Information topic / Information exclusion topic / Questions topic 2 / Question topic 3 / Exclusion question 2 / Exclusion question 3 / Exclusion invite / Invitation / Greeting end

Table 1: Actions categories and listing

Each category includes a variable number of appropriate language and other actions. For instance, the confrontation category contains a considerably larger number of actions than the help section since there is a very limited number of coping behaviours available in dealing with bullying [Woods et al 2003]. The **Help** set

articulates the actions needed to generate offering-help interactions between agents. It covers the interactions needed for the generation of enquiries from agent-to-agent with respect to emotional states and related goals. In addition, this function also generates advice and offers such as help, protection or assistance.

As with the other categories, the Help language and action set category has been designed according to a potential sequential structure. This can be triggered either by an agent asking for the help of another or in response to an aggressive action carried out on a particular agent. The **Confrontation** language and action set provides the necessary content for an altercation between two different agents. This category covers most of the physical bullying expressions and involves threats, insults, orders, aggressive behaviour that leads to aggressive actions and violent behaviour. Finally, the **Socialising** category includes language and actions that can be used in social discussion by pupils in schools (sports, homeworks, music, video games) and language and actions that can be used in generating relational bullying. Relational bullying is different from physical bullying, depending on social exclusion and should therefore be integrated into social interaction, as opposed to help or confrontational actions. Although the structure is simple in theory, its implementation requires a large number of utterances and topics.

3.2 Actions Finite State Machine (FSM)

Each action category also possesses its own organisation and consequently requires the design of its own Finite State Machine (FSM). A language action is coherent to both the system and the user if organised into structured speech sequences. While this has to be taken into account it is also essential that the speech system focuses on organising the possible sequences of utterances and ensure the transfer and communication of content without interfering with the agent action selection mechanism. Since, as with all speech system, there are issues of contextualisation, the utterances that constitute the content of the system are formed of templates that can be filled appropriately by the speech system, based on keyword recognition.

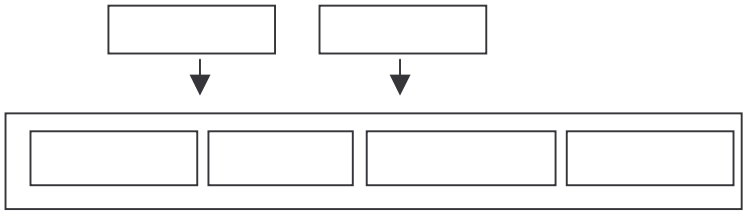


Figure 4: Speech act sequence example

Each FSM integrates the language actions relative to the category itself but also potential elements of answers for discussion or interaction. For instance, the actions ‘DO’ or ‘FORBID’ in a confrontational situation will be followed by the actions

'DENY', 'OBEY', 'ASK WHY' or 'BEG' [Figure 4], to retain conversational coherence.

The VICTEC language actions and utterances have been elaborated according to sequence of actions observed in the scenarios developed by school children mentioned above.

Speech acts are materialised on the FearNot! Demonstrator by utterances. The situation presented in [Figure 4] would produce, in case of denial or obedience from the victim the following exchange [Figure 5].

Speech act	Utterance
DO	You, **[order]** now!
If speech act = DENY	You must be joking, **[rejection] [insult]**
If speechact = OBEY	Ok, but please don't hurt me!

Figure 5: Speech act utterance sequence example

3.3 User-to-agent language action design

Since, the language generated by the user is highly ambiguous and there are no means for the system to understand the meaning of a sentence, the user-to-agent interaction, as we mentioned previously, needs a different approach. As a sentence can only be "understood" by the keywords included in it, it seems sensible to leave the initiative to the agent rather than the user. The fact that the system leads the conversation with the user presents an advantage in terms of believability for the speech system in the sense that, the system can be expectation driven and can expect a certain type of answer from the user and adjust and compare the answer to a set of pre-defined templates. Although the system could not understand its human interlocutor, it could generate a high level of believability and interact with its user by asking simple and adequate questions.

In order for the agent to keep the upper hand in terms of interaction with the child user, it must be the one asking for advice and the one who generally ask questions to which the child user is expected to answer.

It has been decided, due to the high possibility of misspelling from the children who are going to use the system, that the language system includes a keyword recognition feature that should allow it to recognize the intention of the user and make the association with existing categories of actions.

4. Conclusion

This paper has described the interactional structure and articulation of the language system being developed for the VICTEC project and reported on progress made so far. It also detailed the different language actions and their categorisation in relation to the specific theme of bullying.

The language system and its content have been developed based on actual language currently in use amongst school children, however it requires iterative refinement and testing of both its efficiency and the coherence as well as evaluation of its capacity to suspend or limit the initial disbelief commonly generated by this type of system. A series of Wizard of Oz experiments [Maulsby et al 93] along with psychological and usability evaluations [Woods et al 2003] are therefore planned. Further evaluation of the whole FearNot! Demonstrator is also planned: for example, a large sample of children (N: 400) will take part in a psychological evaluation at the University of Hertfordshire in June 2004. However, while the agent architecture of the system and systems integration is still under development, language graphical content has already been produced for preliminary evaluation and the VICTEC team is working with the aim to present a first prototype of the system by April 2004.

References

Austin, J. L. (1962), How to Do Things with Words, Cambridge, Mass.: Harvard University Press.

Bunt, H. (1981), Rules for the interpretation, evaluation and generation of Dialogue acts. In IPO annual progress report 16, pages 99-107, Tilburg University 1981.

Kar2ouche. www.kar2ouche.com Immersive Education

Ortony, A; Clore, G.L. and Collins, A. (1988) The Cognitive Structure of Emotions. Cambridge University Press, 1988

Mauldin, M. (1994), Chatterbots, Tinymuds, And The Turing Test: Entering The Loebner Prize Competition. Proceedings AAAI 94

Maulsby, D., Greenberg, S., and Mander, R. (1993), Prototyping an intelligent agent through Wizard of Oz**.** In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 277-284, Amsterdam, ACM Press.

Searle, J. (1969) Speech Acts. Cambridge University Press, 1969.

Szilas, N (2003) Idtension: A narrative engine for interactive drama: 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE 2003), Darmstadt (Germany) March 24–26 2003.

Weizenbaum, Joseph. (1966) "ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine," Communications of the Association for Computing Machinery 9 (1966): 36-45.

Woods, S; Hall, L; Sobral, D; Dautenhahn, K and Wolke, D (2003): Animated Characters in Bullying Intervention. IVA 2003: Springer-Verlag LNAI 2972 310-314