A mechanism for acting and speaking for Empathic Agents

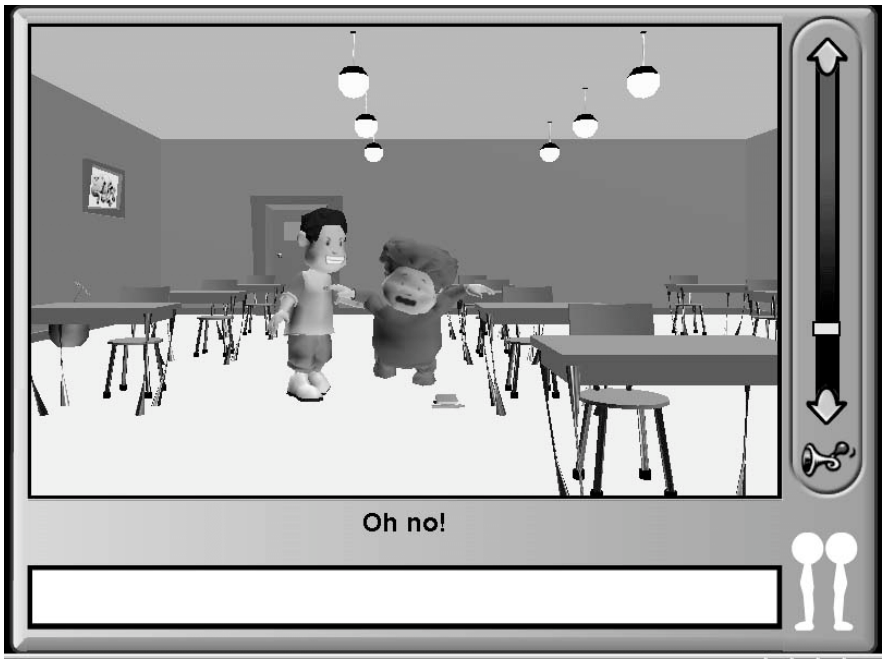
R.S.Aylett, S.Louchart, J.Pickering
CVE, University of Salford

Abstract: The FearNot! application demonstrator, currently being developed for the EU framework V project VICTEC, uses empathic agents as a basic mechanism for engaging children in anti-bullying education. The empathic agents are driven by an affective system in interactional dramas in which both physical actions and language actions are required. This paper 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. The paper also presents early development and implementation work as well as system and speech evaluation planning.

1. Introduction

The EU framework V project VICTEC - Virtual ICT (Information and Communication Technologies) with Empathic Agents - 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 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 (http://www.VICTEC.org) is education against bullying. The project expects to contribute to an understanding of the role of empathy in creating social immersion, and to the evaluation of virtual environment ICT for child users. It also expects to contribute to a deeper understanding of empathy and its role in bullying, and to the relationship between Theory of Mind (TOM) [Woods et al 03] and bullying behavior. The building of empathy between child and character is seen as a way of creating a novel educational experience.

Figure 1: The FearNot ! demonstrator



An output of the project is the FearNot! Demonstrator (see Figure 1), 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 narrative approach undertaken by the VICTEC project is that of Emergent Narrative [Aylett 1999, Louchart & Aylett 03,04]. The research on Emergent Narrative aims at finding and elaborating a narrative structure appropriate and suitable for an optimal use of Virtual Environments, combining the entertainment values of both storytelling and virtual experiencing.

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 [Woods et al 03] 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 Environments (VE), in order to build empathy between the child and the virtual character as the child explores different coping behaviours in bullying.

2. Integrating language and action

Most dialogue systems or talking heads are entirely language-based, with other actions, such as gestures or facial expressions, seen as additions to the main communicative behaviour of a character. In VICTEC however, language interaction is mixed with physical actions which are of equal or greater importance to the language used. 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. These may be accompanied or not by language. Each character involved in the virtual drama episodes of the FearNot! Demonstrator is provided with its own autonomous action selection mechanism with the overall architecture is shown in Figure 2. An appraisal of events and the other characters is carried out, currently using the emotion-modelling system of Ortony, Clore and Collins [Ortony et al 88] and the resulting emotional state is combined with the

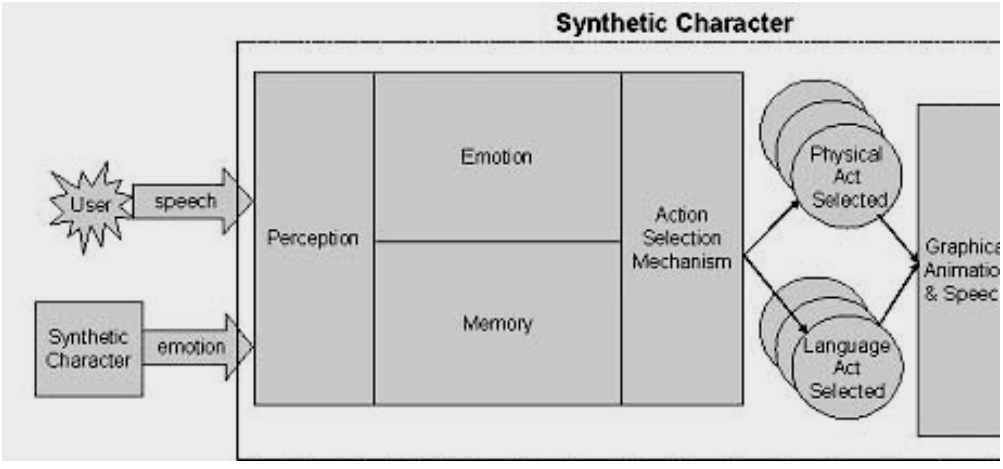


Figure 2 ; Overall agent architecture

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 language 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 03] for other work with this aim). In addition, language and physical 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.

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 the following instances: an object on which the action can be performed (an object in the environment or another character) ; the agent performing the action ; the action priority (used to order and deal with 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, if any, relating to language actions, that should be played, and the animation of the body of the character involved and accompanying gestures. The emotional status of the character feeding into its action-selection mechanism will determine whether the action to be performed is implemented via language action, physical activity or both. Thus depending on its level of aggression, a character may mock or insult another (both language actions), may push it, or may hit it.

Physical actions are realized as animations associated with the graphical character, and its emotional state is reflected by a facial expression, implemented cheaply as a texture change on the character's face. In order to generate the utterance for a selected language action, it has been decided to use a modified shallow-processing approach, based on that originally used in ELIZA [Weizenbaum 66] and more recently in chat bots [Mauldin 94]. Thus the template database typical of such systems has been extended by indexing it via language actions, as discussed in more detail below.

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 the agent-to-agent case, there is no ambiguity

Figure 3 : Agent-child interaction

about the action – whether physical or language – that has been selected by one character to which a second character has to respond. In the case of a language action, there is also no ambiguity about the content of the utterance for which a response must be generated, and since this also comes from the templates database, it is much less demanding than responding to an unlimited repertoire from a human user.

In the child-to-agent case, the problem of dealing with unexpected inputs still exists. Here, the FearNot! demonstrator will specifically drive the conversation by using leading questions with a limited range of options for answer. The question-asking strategy was one used originally by ELIZA, but rather than the more open domain of an individual's emotional problems. FearNot! covers the much more specific one of bullying in the strong context of the episode the child has just seen. We believe that this, together with the provision of 'sentence starters', to help slow typers, will allow the system to behave competently. Wizard of Oz studies are in progress to determine in more detail what language coverage will be required.

FearNot ! draws on techniques and technologies inspired by research in conversational agents [Braun 02,2003; Rist et al 03, Prendinger & Ishizuka 01, 02]. A similar approach has already successfully been implemented in FACADE [Mateas & Stern 03], in which the characters also choose between actions and language when interacting with each other or with a user. However, FACADE's low level of abstraction approach would be hard to manage for VICTEC and would require more development than actual resources allow.

In agent-agent interaction, the language system starts with a 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:

Template:

```
<GreetingWord><Name?><StatusQuestion?>
```

Language Act:

```
<From>Tom</From><To>Luke</To><semanticInfo name="true"
statusQuestion="random"/>
```

Yields one of:

```
Hi Luke
Hello Luke, how are you?
```

Child and character interaction is different. Here the previous language 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. In addition 'sentence starters' (see Figure 3) are provided to help the child with the typing burden and these will provide clues to the language action the child has carried out.

3. The FearNot! Language Action 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 ensuring 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. For instance, in the case of an agent-to-agent communication, the process starts with the selection of a language action and ends with the selection of an utterance. The opposite occurs in the case of agent-to-user communication since the system needs to recognise an utterance via keywords and then select an appropriate language action or action in order to provide an answer to the user.

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, as seen in Table 1.

Category	Content
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
CONFONTATION	Order / Aggressive questioning / Do / Forbid / Defiance / Tease / accusations / Insult / Threat / Aggressive answer / Apology
SOCIALISING	Abandon action / Action / Hit / Lie / Steal / Obey / Deny / Ask why / Beg / Claim back / Leave / Struggle. Greeting start / Topic introduction / Exclusion topic introduction / Information topic / Information exclusion topic / Question topic 2 / Question topic 3 / Exclusion question 2 / Exclusion question 3 / Exclusion invite / Invitation / Greeting end

Table 1 : Language action categories

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 category 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)}

A language action is coherent to both the system and the user if organised into structured speech sequences. However it is also essential that the language system focuses on organising the possible sequences of utterances without undermining the agent action selection mechanism. A Finite State machine (FSM) approach is taken to limit the number of language actions from which the action-selection mechanism can choose. Each action category possesses its own organisation and consequently requires the design of its own FSM.

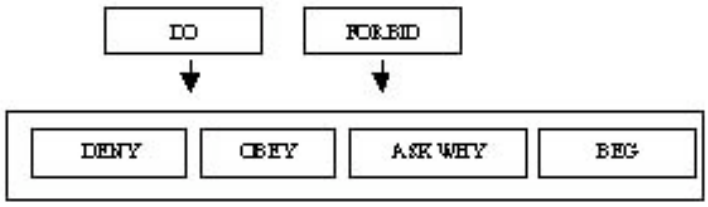


Figure 4 : An example of Language Action sequencing

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' - Figure4 - to retain conversational coherence. These language actions and utterances have been developed by analyzing the scenarios developed by school children in Kar2touche already mentioned.

Language acts manifest themselves in the FearNot! Demonstrator through utterances. The situation presented in Figure 4 would produce, in the case of denial or obedience from the victim the following exchange shown in Table 2.

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

Table 2: An example sequence of speech act utterances

3.3 User-to-agent language action design}

Since the language generated by the user may be ambiguous and there are no means for the system to understand the meaning of a sentence, the user-to-agent interaction, needs a somewhat different approach. As a sentence can only be "understood" via matching on the keywords included in it, it is highly desirable to keep the conversational initiative with 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 language system in that it can be expectation driven, anticipating a certain type of answer from the user and comparing this answer with a set of pre-defined templates. Although the system cannot not understand its human interlocutor, it can, we believe, generate a high level of believability by asking simple and adequate questions to which the child user is expected to reply.

This type of keyword recognition should also allow it to cope with a certain amount of misspelling, recognizing the intention of the user and making the association with existing categories of actions. These categories divide into two - the agent language actions and utterances and the user's answers, seen in Table 3.

Category	Actions
AGENT	Ask for advice / Ask again / Prompt / Cannot understand statement /Ask for reason / justification / Thank user for advice / Confirm advice with user /Express reproach to user / advice rejection / Express disappointment towards user /Report result of interaction /Beg for help
USER	Give advice /Refuse to give advice /Ignore the agent / No answer / No helpful comments /Advice confirmation / Justification

Table 3 : User and Agent Language actions lists

4. Implementing Language Acts

As discussed above, Speech Act theory has been taken as the basis for VICTEC's Language Actions. However there are well-known problems with Speech Act theory which must be confronted in implementing Language Actions. The problems are: speech acts alone contain no semantic information; speech acts are not unique; and speech acts cannot in general be mapped to syntax. It is claimed [Jurafsky & Martin 00] that classifying utterances into speech acts is an AI complete problem, meaning that a human being, or a computer system equivalent to a human being, would be required to correctly classify them.

4.1 Microgrammars

Although the general problem of classifying speech acts is currently non-formalisable, it is possible to produce automatic classifiers that give partial coverage for common

acts. The method for doing this exploits the fact that many speech acts correlate to structural features in a conversation. These structures, introduced by [Goodwin 96] have been called *microgrammars.* They comprise sets of features which are classified into three different types.

Words and collocation: certain words and particularly combinations of words (collocation) indicate some speech acts. For example the words `who, when, where', indicate questions.

Prosody: the tone of voice used in an utterance may indicate its intended act. In English, questions, for example, can be indicated by a rising intonation at the end of a sentence.

Conversational Structure: the current context and the immediate predecessor statements may give an indication of the speech act. A simple example of which is that the utterance after a question is probably a reply or a request for clarification.

In the case of a textural system such as that currently being used in FearNot!, prosody will have no part to play. The burden of the work will have to be achieved using pattern matching to identify words and collocation. Hopefully, some support can also be provided through the use of context.

4.2 Language Actions

Because the VICTEC project is centred on the development of autonomous agents that interact in a virtual environment by the use of actions, it was natural to use speech acts to define the agent's language system. This would allow the agent to remain in an action reception, action appraisal, action selection loop. The problem is the lack of semantics and multiple definitions of speech acts. To allow for the first, some semantic information has had to be added to the agents actions. We have called the combination of speech act plus semantic information *language actions*.

We intend to solve the second problem of how to identify sentences with speech acts by applying microgrammars to the very small set of sentences that have been classified in the knowledge base. A microgrammar can be written for each speech act. When parsed in conjunction with the semantic information and contextual knowledge of the source and sender of the speech act the microgrammar will generate a sentence. For example consider the act of greeting a person. The set of possible sentences is very small, consisting of a greeting word, possibly the name of the person being greeted, and possibly a greeting question.

```
Hello
Hello Sue
Hi
Hi Tom
Hi Jo, how are you?
```

We can immediately see a general form to these sentences which can be expressed in Backus Naur Form (BNF) as follows:

```
<Greeting word> <ToName><status_question>
<Greeting word>    = <Hello>
                   = <Hi>
<ToName>           = <receiver>
<status_question>  = <how are you>
                   = <are you all right>
```

Here, the term <*receiver*>is a context variable that is set by the semantic information in the language act.

4.3 Implementing Agent to Agent Language Actions

The database of language acts will be specified in XML, which has the expressive power of a context free grammar and so can express BNF statements. Each language action would be implemented as a template consisting of rules. The rules are implemented separately and may be recursive.

A language action generated by an agent will contain the name of the sending and receiving agents, the type of action and some semantic information, that is action specific. When received, the template for the requested action is found, context variables are set and then the rules are repeatedly applied until a response has been formed. For example the following XML specifies a request for a greeting from agent Tom to agent Sue, with an optional question about Sue's current status.

```
<Type> Greeting </Type>
<Sender> Tom </Sender>
<Receiver> Sue </Receiver>
<SemanticValue name= “true” statusQuestion= “random”\>
```

Fist the variables `sender` and `receiver` would be instantiated to the names Tom and Sue, then the template for the language act looked up. Using a greeting language act as specified in 4.3 the template requires a greeting word that can be `Hello' or `Hi'; as there is no other information a random choice would be made. Next, because the name attribute of the semantic information was set to `true` the `ToName` must be added. The rule for this evaluates to the context variable `receiver`, so the value of Sue would be added to the reply.

Finally, the status question attribute of the semantic information was set to `random` so the the language system will chose with equal probability between tadding and not adding a status question. If a question is to be added the rule is evaluated which gives a random choice of two possible questions. This results in one of the six following possible greetings being generated:

```
Hi Sue
Hello Sue
Hi Sue how are you?
Hello Sue how are you?
Hi Sue are you all right?
Hello Sue are you all right?
```

The rules can be recursive, allowing a rule to contain other rules, which will allow the case and gender agreements of German and Portuguese to be applied.

4.4 Implementing User Agent Dialog

In the case of user agent dialog the problem of classifying the user's speech acts and extracting the semantic information for appraisal by an agent must be addressed. As was stated above, this is in general a very difficult problem. However the problem may be simplified by noting three features that will apply in the case of FearNot!

- The dialog will be very short and focused only on the previous bullying
- The users will be children of age 10 and so will only type simple sentences.
- In order to help the children buttons providing part formed sentences will be provided.

It is hoped that these features will so constrain the input domain as to allow the identification of speech acts using pattern matching to look for words and collocation supported by some conversational structure information.

5. Conclusions and future work

In this paper we have discussed the design of a language system for the VICTEC project which integrates language with physical action, and can be used both between synthetic characters in virtual dramas and between a character and a child user. The bullying-specific content of language actions has been described and the issues involved in implementing the system, based on the concept of speech acts, also considered. Wizard of Oz trials are currently being organised to help to validate the design.

The language system is now being implemented and will shortly be integrated with the overall FearNot! demonstrator. A version in German will then be immediately developed for an evaluation exercise to be run with German school children. The English version of the system will be used in a large-scale trial with 400 children to be run in June 2004.

The most desirable extension to this system is seen as the incorporation of speech input and output to avoid the typing overhead for the target user population. However further issues will have to be confronted here: firstly, an extra level of ambiguity added into the user's input, and secondly, the deficiencies of text-to-speech, with their impact on overall believability and the development of empathy between child and victimized character.

References

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

Aylett, R. (1999) Narrative in virtual environments – towards emergent narrative. 1999; AAAI Symposium on Narrative Intelligence pp 83-86

Aylett R, Louchart S. (2003) Towards a narrative theory of VR. Virtual Reality Journal, special issue on storytelling. Virtual Reality Journal Volume 7 January 2004

Braun, No (2001) Storytelling & conversation to improve the fun factor in software applications in: Mark A. Blythe, Andrew F. Monk, Kees Overbeeke, and Peter C. Wright (ed.): Funology, From Usability to Enjoyment , Chapter 19, Kluwer Academic Publishers, Dordrecht, ISBN 1-4020-1252-7, April 2003

Braun, N. (2002) Automated Narration - the Path to Interactive Storytelling In: Workshop on Narrative and Interactive Learning Environments, Edinburgh, Scotland, 2002

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.

Goodwin, C. (1996) Transparent Vision. In: Interaction and Grammer, eds E Ochs; E Schlegloff & S Thompson,Cambridge University Press, 1996

Jurafsky, D & Martin, J (2000) Speech and Langugage Processing, Prentice Hall, 2000

Kar2ouche. www.kar2ouche.com Immersive Education

Louchart S, Aylett R (2002) Narrative theories and emergent interactive narrative; Proceedings Narrative and Learning Environments Conference NILE02 Edinburgh, Scotland pp 1-8

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

Mateas, M & Stern, A. (2003) Integrating Plot, Character and Natural LanguageProcessing in the Interactive Drama Façade. Technologies for Interactive Digital Storytelling and Entertainment (TIDSE) conference, Proceedings 139-152

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.

Prendinger, H. & Ishizuka, M. (2001) Let's talk! Socially intelligent agents for language conversation training. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 31, Nr. 5, 2001, pages 465-471.

Rist, T;Schmitt, M; Pelachaud,, C. & Bilvi, B. (2003) Towards a Simulation of Conversations with Expressive Embodied Speakers and Listeners. CASA 2003: 5-10
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