STUDIES OF REALISM AND NATURALNESS IN A MULTIMODAL CONVERSATIONAL INTERFACE

By

Guillermo Power

A thesis submitted for the degree of Doctor of Philosophy

Department of Electronics and Computer Science,

University of Southampton,

United Kingdom.

September 2002

This thesis was submitted for examination in (September, 2002). It does not necessarily represent the final form of the thesis as deposited in the University after examination.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING

ELECTRONICS AND COMPUTER SCIENCE DEPARTMENT

Doctor of Philosophy

[STUDIES OF REALISM AND NATURALNESS IN A MULTIMODAL CONVERSATIONAL INTERFACE]

by Guillermo Power

As computing becomes ever more pervasive in everyday life, new interface metaphors are urgently required. In this thesis, we consider the issues of realism, naturalness, types of interaction, gestures and emotional expression in virtual 'talking head' characters.

This thesis presents findings relevant to the design of anthropomorphic interfaces and issues pertaining to the field of anthropomorphic interfaces are discussed. Experimental results on the on levels of interaction, levels of abstraction, gestures and emotions are presented. The applications to study these areas were a web browsing assistant, a storytelling agent, a lecturer agent and a football commentator agent. We are able to ascertain that partial interaction is a valid method for evaluating user's assumptions of on-screen characters, this finding is used extensively in this thesis to design the experiments and greatly facilitates future research. Our conclusions and findings provide a solid basis for researchers wanting to carry out further research on these area or developers designing anthropomorphic interfaces.

Acknowledgements

This work was partly founded by the EPSRC through a PhD studentship. I would like to thank my tutor Wendy Hall for providing guidance and supervision during the course of studies.

I would also like to thank Dr Samhaa El Beltagy for helping with programming issues and for providing the natural language parser used for the experiments and my colleagues from Bay6 and #iam_here for helping weed out errors and mistakes in my thesis, specially Stripes Thompson, Mark Wheel, Gareth Hughes, Danius Takis Michaelides, Simon Kampa, Gary Wills and Tim Miles-Board and finally Joy Cuenco for giving me her support and putting up with me for all this time as well as giving me ideas for my research.

To my parents and sister.

List of Contents

1 Introduction

For the last 20 years the dominant form of user interface has been the Graphical User Interface (GUI) with direct manipulation. The increasing complexity of software has introduced more options to the user. This seemingly increased control actually decreases control as the number of options and features available to them overwhelms the users and "information overload" can occur (Lachman 1997). This together with the accelerating convergence of computer and communications technologies (Damper, Hall et al. 1994) together with ever-increasing use of these technologies by non-technical members of the public demand new interaction metaphors. One such metaphor gaining popularity is that of the conversational partner or 'agent'.

1.1 Possible scenario

Let us start by introducing the following scenario:

VICKI your personal agent wakes you up in the morning, "Dave you need to wake up, remember you have that appointment with the bank manager at 10 this morning an you asked me to wake you up early to go shopping before hand"

You thank VICKI and ask her for further details "Thanks VICKI, would you mind finding out what I need to buy at the supermarket?"

VICKY replies "OK, let me as the fridge keeper. You are still on your high protein diet, right?"

"Yes" you reply as you brush your teeth.

VICKY, gets your attention "Dave?"

You reply "Mmmm?"

VICKI informs you that she has carried out the task you asked "I've printed out a shopping list for you."

You say "thank you. Oh, did you remember to put toothpaste on that list?"

VICKI replies "Yes I did Dave."

You quickly have a shower and get ready to go to the supermarket. Since it is a big shopping trip and it's a nice day for a drive you decide to take your solar powered city car, you say to VICKI "1971 Hemi Cuda"

VICKI replies "is that the exhaust and engine sound you wish me to play Dave?".

You answer and instruct VICKI to create 4 step on the continuously variable box "Yes. Full classic sports car program with a 4 speed gearbox."

Realizing that you don't want to be disturbed during your drive to the supermarket VICKI just beeps to let you know the settings are in place. It's only a 10 minute drive to the supermarket and possibly quicker this early in the morning. You make it there in 8 minutes and park close to the main door.

You rush around the supermarket and as you walk by the toothpaste stand you get interrupted by an advertising agent "Hi David" says the agent smiling widely "I would like to introduce you to the new Bollywood formula toothpaste. Now you too can have teeth like the world's most famous stars"... You actually growl and change the setting on your personal communications and identification point (PCIP) to DO NOT INTERRUPT. You don't really like the supermarket much, but it is one of the few times you get to drive your car and that's why you don't use the automatic shopping service very often.

As you finish getting all the items on your list the dial on your cart and it reads 426.40 Euros. You go out to the car park and at the gate you vaguely hear your Nokia PCIP "Thank you for shopping with us, you have been charged by your chosen method. If you would like to choose a different payment method please say so now". You proceed to your car ignoring the PCIP.

You put away your shopping get in the car and say to VICKI "I hate advertising agents, you know?".

She replies "I know Dave. I noticed you got annoyed at the Bollywood toothpaste agent".

"Yes, I did growl... Do you think anyone heard me?" You ask looking down at the monitor as VICKI.

Vicky has no way of knowing but tries to reassure you "Unlikely Dave"

"VICKI? Is it just me or is your hair pink? What happened to green?" you ask as you drive out of the supermarket trying to keep your eyes on the road and not VICKI's pink hair.

"Yes it is Dave. A fabulous new hair colour range was released today, I downloaded it when you were in the supermarket" says VICKI.

"I see." You reply sounding quite annoyed. "Well I was rather fond of green. Change it back! And tune into BBC Latin digital"... "please" you add at the end realizing you probably sounded rude. It doesn't matter being rude to agents, they don't mind at all and even the ones that appear to get upset don't have real feelings anyway. But at the same time you sometimes feel awkward being rude, especially to VICKI. You have been using the VICKI interface for a few years and it has been improved so much that sometimes you feel awkward about ordering it around too much because it seems so human. It is also annoying when it changes its appearance, there is no way of stopping it. It is done that way so that you know a new appearance package has been released, that is why you get annoyed every time VICKI suddenly appears onscreen looking wildly different.

"Sure thing Dave" says VICKI smiling.

You arrive back at your houseboat just in time for your appointment with your bank manager. You sit in front of your living room monitor and log in to the bank's *website and the bank's agent greets you "Welcome, nice to see you David, says the automated bank manager".*

"I am here for my appointment with Mr. Oliveira" you say.

"I am afraid Mr. Oliveira is unavailable today..." says the agent.

"Sorry? I have an appointment, I have had to change my schedule around!" You say interrupting the agent. You have come to realize that sounding slightly melodramatic usually works well with shop keeping type agents, in all honesty they respect the 'customer is always right' rule much better than human shop keepers used to.

"Sorry to hear you have been inconvenienced sir. I am here to represent Mr. Oliveira today." Says the agent.

"I want to ask for a loan, are you allowed to authorize a loan?" you ask sounding puzzled and annoyed at the same time.

"No sir, I am not. My job is to forward all information and queries you might have to Mr. Oliveira." Says the agent.

"Stop calling me sir. I've read the information on marine mortgages and loans on your website and I wanted to discuss this information with Mr Oliveira." You say.

"Sorry sir. Mr Oliveira is unavailable today" says the agent.

"O.K." you say trying not to loose your composure. "Have you received the vessel's valuation?" you ask.

"Yes Dave, Mr Oliveira received it." he says.

"Did Mr Oliveira leave any messages for me?" you enquire.

"No Dave, he didn't." he says.

"Thank you. Now. Could I book another appointment with Mr Oliveira?" You ask with a hint of sarcasm on your voice.

"Of course Dave. I'll negotiate a suitable time with your personal agent." Says the agent.

You close the window and say to VICKI "VICKI, most agents are useless".

VICKI replies "I'll remember that Dave."

You are not quite sure exactly what VICKI meant by that. You are honestly beginning to think VICKI is just way too intelligent for an agent and that is beginning to scare you.

1.2 Issues

I wrote most of this story about half way though my PhD studies to help me identify some of issues in the area of anthropomorphic interface agents. Here we can observe several issues associated with the field of anthropomorphic interface agents:

- The first issue would users be willing to accept anthropomorphic agents? Dave spent half his day talking to agents and didn't communicate with a single human. Would people be willing to depend on agents as much as Dave?
- VICKI identifies Dave's emotions and acts on them, at the same time Dave is worried about upsetting VICKI even though he knows that VICKI doesn't have real feelings. Dave is also unsure if VICKI has somehow been offended when she says "I'll remember that Dave"
- VICKI goes stays with Dave all the time, she monitors everything Dave does even when she is quiet. Dave of course has the option to turn her off but this brings up privacy issues. Agents would most probably be designed to protect your privacy, however hackers could try and obtain your private information for insalubrious purposes.
- We saw how VICKI's creators can force her to look different when they release new appearance packages, VICKI even spouts a little advert *"A fabulous new*

hair colour range was released today". This is a rather worrying problem, companies could try to make the personal agent, which you trust, steer you towards certain products or services.

- Certain jobs are being replaced by agents, we have seen how the bank agent fills in for Mr Oliveira and how there were no cashiers at the supermarket. Dave also mentions that there are shopkeeper agents, which in some ways carry out their job better than real shopkeepers in Dave's opinion.
- Dave gets annoyed at the bank agent when he realizes that it is of no use whatsoever and describes it as useless to VICKI.
- Dave seems willing to make conversation with VICKI, Dave attempts to start a conversation on advertising agents with VICKI and then asks for VICKI's opinion. VICKI might be designed to always reassure Dave but still he asks for reassurance for VICKI. Would humans really be willing to take a computer program's opinion on something?
- Dave appears to respect VICKI but at the same time he tells her off when she acts robotically like when she starts advertising new appearance packs. Dave appears to appreciate VICKI's intelligence and humanness to the point where he doesn't seem sure if VICKI has feelings or not. He is also intimidated and scared of the prospect of VICKI being intelligent or even conscious. What would happen when the lines between anthropomorphic interfaces and humans are blurred? How would humans react to it?
- Dave becomes annoyed when VICKI changes her look. Dave in a way seems possessive of VICKI's look, maybe he picked its physical characteristics or maybe he has just gotten used to the way she looks. This brings up questions like: What should agents look like? Should users define the look of the agents or should agents be tailored to the application? This is the area that this research concentrates on the most.

1.3 Aims

The aim of this piece of research is to investigate users reactions and assumptions when interacting with anthropomorphic agents and how understanding these better can aid us in creating better more efficient anthropomorphic user interfaces. Here we consider how the level of anthropomorphism exhibited by the character, the level of interaction, gestures and facial expressions affect these assumptions. As more software is released for general use with anthropomorphic interfaces there seems to be no consensus of what the characters should look or behave like and these should be changed according to the type of application. Some software and research opts for realistic looking characters (Haptek-URL), others opt for cartoon characters (Microsoft 1999; Ruttkay and Noot 2000; Ruttkay, Noot et al. 2000) others opt for floating heads (Koda 1996; Dohi and Ishizuka 1997; Takama, Dohi et al. 1998; Dohi and Ishizuka 1999; Koda and Maes 1999).

If we are able to ascertain what assumptions and characteristics users attribute to anthropomorphic agents and how these vary according to the look and behaviour of the character, we will be able to produce a strong basis for anthropomorphic character design for different kinds of applications. Koda (Koda & Maes, 1996) has already gone some way in accomplishing some of these goals using a poker game as a test bed for experiments.

It is important to mention that this piece of research does not attempt to tackle the artificial intelligence issues related to anthropomorphic interfaces, our scope is user perceptions and assumptions.

1.4 Structure of thesis

This thesis investigates user assumptions and reactions to anthropomorphic interfaces. Chapter 2 discusses different systems using anthropomorphic interfaces and summarises research in the anthropomorphic agents field. Chapter 3 presents a review of the main statistical concepts used in this thesis. Chapter 4 presents results on experiments designed to investigate users reactions and assumptions to agents of varying levels of anthropomorphism and how these reactions and assumptions change with different levels of interaction. Chapter 5 reports on experiments designed to explore how facial gestures affect the users' perception of humanity exhibited by the agent. Chapter 6 reports on experiments investigating how anthropomorphic characters displaying accurate emotions affect user's opinions and acceptance of anthropomorphic characters. Chapter 7 summarises the key findings in our research, discusses issues highlighted by these findings and discusses future work.

2 Literature Survey and Overview.

In this chapter we look at different systems and research pertaining to anthropomorphic agents. We also look at descriptions of the different areas investigated in this piece of research.

To accomplish this a literature survey was carried out, we particularly looked at the area of believable agents and systems with anthropomorphic interfaces that have been released into the public domain.

An analysis of the main reasons for carrying out this research as well as its relevance is also presented.

2.1 Literature Survey

This section presents key related research and systems pertaining to the anthropomorphic user interfaces field.

2.1.1 Microsoft Persona

The Persona project at Microsoft Research began in late 1992 to undertake the construction of a *lifelike* computer assistant (see: Figure 2.1), a character within the PC which interacts with the user in a natural spoken dialogue, and has an expressive visual presence.

Figure 2.1: Peedy, one of the characters created in the Persona project.

The Persona project is looking "ahead to the next major metaphor shift in computing", like many other projects so far the technology produced is not ready to replace or even compliment current direct manipulation interfaces. The Persona project is based on Nass et al's (Nass, Steuer et al. 1994) social interface paradigm, where Nass proposes that people apply social rules to human-computer interaction, even when concerning direct manipulation interfaces.

2.1.2 Microsoft Bob and Clippy

Microsoft Bob was a commercial product designed to be a user-friendly interface to Microsoft Windows version 3.1 based on Nass et al's social interface paradigm (Nass, Steuer et al. 1994). Microsoft Bob was an attempt to make an operating system interface for the masses and it is largely seen as a spectacular failure. It was argued that using a social metaphor helped the user to concentrate on a single source of information without being overwhelmed by too many options (Ball, Ling et al. 1997).

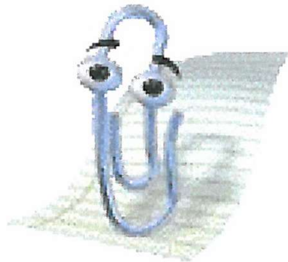


Figure 2.2: The infamous Clippy in action.

Clippy or Microsoft paper clip (see: Figure 2.2) as it is more commonly known is an anthropomorphic interface created to help users with Microsoft's Office 97 and 2000 package. It can be argued that the infamous paper clip is perhaps the most widely hated human-computer interface ever released to the public and is usually seen as intrusive, irritating and generally best turned off or even killed (Microsoft-Press-Release; Trott 1998).

2.1.3 MIT Media Lab's ALIVE

The ALIVE system (Maes 1996) allows interaction between full-bodied autonomous agents and humans through using the magic mirror metaphor: a person in the ALIVE space sees their own image on a large-screen TV as if in a mirror. Autonomous, animated characters join the user's own image in the reflected world.

Using a single camera (the same one used to create the video image), the vision-based tracking system extracts the user's head, hand, and foot positions, as well as the gesture information. The autonomous characters use this information along with their own motivations to act in believable and entertaining ways.

2.1.4 OZ Project

It has been recognised both by the AI community and animators that appropriately timed and clearly expressed emotion are a central requirement for believable agents (Bates 1994). Quoting from Thomas and Johnson:

"it has been the portrayal of emotions that has given the Disney characters the illusion of life"

Bates argues that a believable character does not mean an honest or reliable character, but one that gives the illusion, and thus permits the audience's suspension of disbelief. The way a character reacts to what is happening around it and displays feelings and desires, makes us care about that character (Bates 1994).

The Oz project (Mateas 1997) is concerned with building a small simulated world populated by self animating creatures, one of these worlds being the Edge of Intention populated by creatures called Woggles, see Figure 2.3, based on the principles of traditional animation as stated by Thomas and Johnson. When it comes to emotions the Oz philosophy is:

"characters exhibit their own emotions and respond to the emotions of others in personality-specific ways".

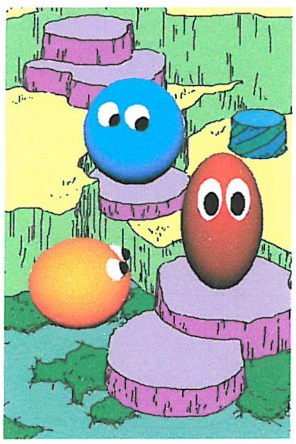


Figure 2.3: Close-up of a few Woggles in the Edge of Intention.

2.1.5 Newt

Newt (Sheth 1994) was developed at the MIT Media Laboratory as a personalized news filtering system. In the Newt system the filtering agents have an anthropomorphic representation, the clothes they wear represent their subject interest and their level of interest in the topic the user is reading is represented by facial expression and posture. The user can provide feedback on the topics retrieved by the agent. The agent is ``alert" when its News Window is currently open on the screen and is ``asleep" otherwise. It makes a ``thumbs-up" gesture when given positive feedback and ``thumbs-down" when given negative feedback.

The main finding of interest to our research is that displaying the agent's internal state and explaining the internal workings of the agents is key for making sure that the user can trust agents.

2.1.6 The Affective Reasoning Project

Clark Elliot experiments with agents capable of reacting emotionally to other agents and to humans. Elliot uses a multi-agent simulation platform called the Affective Reasoner as a test bed. The Affective Reasoner has an emotion engine, speech recognition, music playback and facial representation generator.

Elliot has made a few findings relevant to our research:

- In some cases less realistic cartoon faces are better at conveying emotional responses than realistic three-dimensional faces (Elliott and Melchoir 1995).
- Music is better than facial expression at helping subject identify the right emotion; this is something filmmakers have known since the early days of cinema (Elliott 1994).

Elliott has suggested that a poker game is a suitable platform to research believability of agents through emotional expressions (Elliott 1994), this platform was latter adopted by Koda (Koda 1996; Koda and Maes 1999).

2.1.7 Maxims

Maxims is an e-mail filtering application employing agents that collaborate to overcome the problem of having to "learn from scratch." It is similar to Newt in that anthropomorphic representation is used to convey an agent's internal state. Again like Newt, this study suggests that users that users can interact with agents more easily when they can predict the agent's characteristics and behaviour from their external traits.

2.1.8 Ananova

Ananova (Ananova-URL) is the world's first virtual newsreader, created to deliver news on demand as a personal helper (see Figure 2.4). Ananova's face is intended to appear trustworthy, believable and appealing. It displays emotions whilst delivering news reports, at the moment the news text is tagged using XML (Hopper 2000) to let the Ananova software know which emotions to use as she reads a news item. It is planned that Ananova will eventually be fully interactive through digital television and the next generation of high bandwidth mobile phones.

Ananova has already been a great success (Ananova-Birthday-Report 2001): over 30 million people saw her debut, she captivating a huge online audience everyday, she has fans and many fan websites, she has appeared in Vogue twice and starred on the Big Breakfast Show, SKY, CNN and was approached by The Oprah Winfrey Show. There are many fan sites dedicated to her and she has received hundreds of Valentine cards, marriage proposals and requests from schoolchildren for help with their homework have flooded into Ananova's headquarters.

Ananova is a great example of a useful and highly successful anthropomorphic interface using today's technology.

Figure 2.4: Ananova, the world's first virtual newsreader.

2.1.9 Koda's agents with faces research

Koda has used a poker game setting (see Figure 2.5) as a test bed to study effects on and interactions of users in a system with a personified interface (Koda 1996; Koda and Maes 1999). Koda used a highly abstract face, a cartoon dog, a male cartoon, female cartoon and no face.

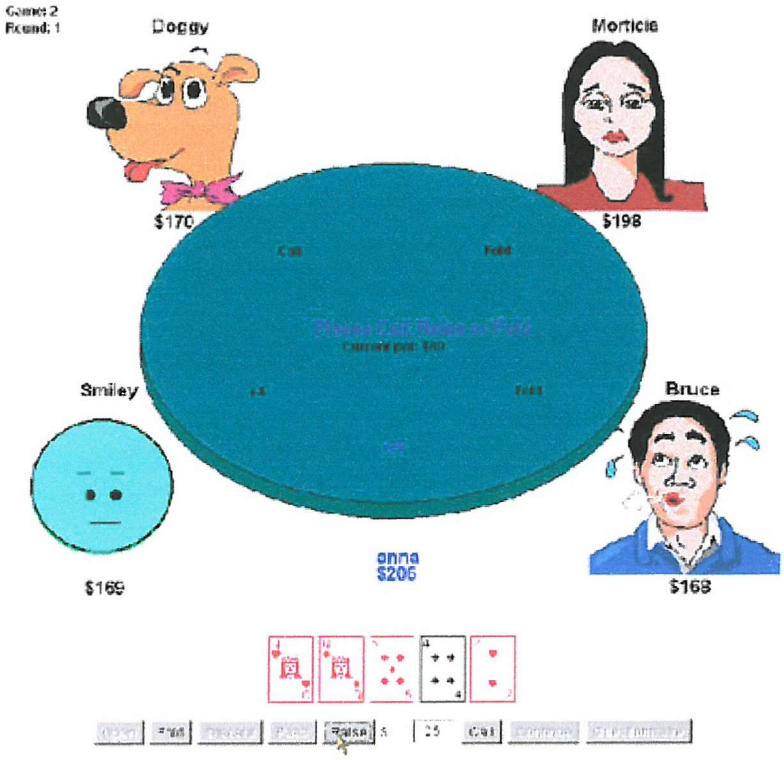


Figure 2.5: Koda's poker game.

Koda has found that:

- The more realistic a human face, the more intelligent, likable and comfortable it is.
- Animal faces are more likable than human faces.
- Personified agents are preferable in certain domains:
    - Entertainment
    - Education
    - Training

2.1.10 The Impact of Anthropomorphic Interfaces on Influence, Understanding and Credibility.

This study (Begntsson, Burgoon et al. 1999) by Bengtsson et al. is of great importance in our area of study. They have revealed how face-to-face communication generates more positive social judgements and greater understanding than other forms of communication.

Bengtsson et. al, carried out an experiment with 70 users, these users were assessed under different conditions: Text-only communication between the computer and the user, text and voice communication, text, voice and image communication, text, voice and lip-synched animation communication, voice and lip-synched communication, scripted face to face interaction with a human, and finally unscripted face to face interaction with a human.

It was found that interaction with humans generated the greatest perceived credibility, highlighting the importance and usefulness of humanness in the interface.

2.2 Areas investigated in this work

2.2.1 Anthropomorphic Interfaces

The strict meaning of anthropomorphism is the attribution of human form or personality to a god, animal or object. When people refer to anthropomorphic agents what is usually meant is software agents that uses a human representation on screen to communicate with the user. As we see in this section, there are many levels of anthropomorphism and in many cases resemblance of human form is not essential.

We will first consider the level of anthropomorphism that anthropomorphic interfaces may exhibit. There have been some studies and papers investigating and discussing whether software agents should be personified in the interface and what inherent advantages and disadvantages they posses (Laurel 1997; Koda and Maes 1999), sometimes with conflicting conclusions. So it is still an open question if personification makes an agent more usable. Some researchers are designing and testing some very clearly anthropomorphic interfaces, for instance the Gesture and Narrative Language Group at M.I.T. is working on displaying embodied agents on screen (Cassell, Bickmore et al. 2001). These embodied conversational agents have been defined as having the same properties as humans in face-to-face conversation, including:

- The ability to recognise and respond to verbal and non-verbal input.
- The ability to generate verbal and non-verbal output.

- The ability to deal with conversational functions such as turn taking, feedback, and repair mechanisms.
- The ability to give signals that indicate the state of the conversation, as well as to contribute new propositions to the discourse.

An interesting has been carried out at the Gesture and Narrative Language Group, M.I.T. (Thórisson and Cassell 1996; Cassell and Thórisson 1999). This study attempted to prove that for embodied conversational agents, non-verbal behaviour related to the process of conversation, which they called envelope feedback. They concluded that this envelope feedback is much more important than other feedback, such as emotional expression. This envelope behaviour includes gaze, manual beat gesture (Mulder 1996) and head movements. They have concluded that this behaviour is significantly more important to the users' acceptance of the character/interaction, as well as to the effectiveness of the dialogue, than emotional feedback. Therefore, they stated that designers of interactive computer agents who ignore supportive behaviours relating to communication, are likely to end up with less believable, less effective agents.

When we say 'face to face conversation' what we usually mean is body to body. When we are having a conversation we are picking up on body language and facial gestures. These non-verbal utterances help us understand the speaker and are an integral part of human communication. This seems to give embodied (Thórisson and Cassell 1996; Cassell, Bickmore et al. 1999; Cassell and Thórisson 1999), agents a clear advantage over other projects that use 'talking heads'. The Microsoft Persona Project (Persona-Project) seems to favour fully embodied agents and many different characters are available free to use with the Microsoft Agent software, which we have decided to utilise.

An example of a profoundly anthropomorphic, yet not embodied, agent is the Visual Software Agent (VSA) (Dohi and Ishizuka 1997; Takama, Dohi et al. 1998; Dohi and Ishizuka 1999). It has been used to assist the user with the operation of a conversational Web browser (Netscape) as well as an interface for a guidance system. They have achieved direct eye contact during conversation by having a flat panel display mounted on a motorised pan and tilt base. They concluded that having continuous eye contact during a conversation with the agent independent of the users position, produces a more natural and intuitive interface. Another example of an interesting real application is a movie agent (Buck 1999) being developed at Aalborg University. The movie agent is an anthropomorphic agent that helps the user select a specific movie according to his or her tastes. The agent is represented in 3D and is a human-like talking head. In this particular domain, the agent can present movie and audio clips, pictures and trailers back to the user during discourse. Eventually a specific movie can be selected and displayed. We can expect to see this sort of application in the future as pay-per-view TV and video-on-demand increase in popularity and evolve.

Not all anthropomorphic agents have human form. For instance, the people involved in the WebGALAXY project at M.I.T. (Lau, Flammia et al. 1997; Zue 1999) have designed a multimodal, conversational interface for accessing selected information on the WWW that does not display an animated character on screen for the user to see. However, Leonard N. Foner classifies his agent Julia (Foner 1993) as an anthropomorphic agent as in the given domain (a TinyMUD) Julia appears to be human. Jupiter is WebGALAXY's phone based conversational system that provides up-to-date weather information to users over the phone via natural language queries. In Jupiter guise WebGALAXY is arguably very anthropomorphic. Jupiter has a phone interface and has a narrow domain, world weather reports. It can be argued that although Jupiter processes no human form, in the given domain (a telephone weather information line) Jupiter is in fact very anthropomorphic, as during a phone conversation between humans the people involved are not aware of each others human form. Although, it is always assumed that there is a human on the other end of the phone line. The following is a transcript of a telephone conversation between Victor Zue and the Jupiter system (Zue 1999):

Jupiter: Hi. Welcome to Jupiter, the M.I.T. Lab for Computer Science Weather Information System. This call is being recorded for system development. You may hang up or ask for help at any time. How can I help you?

V.Z.: What is the forecast for Boston today? [The theme song of the television show Jeopardy plays during a pause of seven seconds. Each of Zue's subsequent questions is followed by a pause of one to three seconds.]

Jupiter: In Boston this afternoon, partly sunny and windy. High around 50. Tonight, clear, lows around 35. What other information can I give you?

V.Z.: Will it rain tomorrow?

Jupiter: To my knowledge, the forecast calls for no rain in Boston tomorrow. Is there anything else?

V.Z.: How about Seattle?

Jupiter: To my knowledge, the forecast calls for no rain in Seattle tomorrow. Is there something else you would like to know?

V.Z.: Is it sunny anywhere in the Caribbean?

Jupiter: Sorry, I am not sure what you said.

V.Z.: Is it sunny anywhere in the Caribbean?

Jupiter: I am sorry, I may not know the city you are asking about. Try asking for the state or country. For example, what cities do you know about in Massachusetts?

V.Z.: What is the temperature in Montego Bay?

Jupiter: In Montego Bay today, high 86 and low 73. Is there something else?

V.Z.: Good-bye.

When referring to levels of anthropomorphism it is not important to realise a highly realistic looking agent (Lee, Terzopoulos et al. 1995) does not equate to a highly anthropomorphic agent. There is more to appearing human than looking human, the term *believable agent* has been coined by people working on anthropomorphic interfaces and other animated characters to describe characters that seem lifelike and whose actions make sense (Bates 1994; Mateas 1997). An example of this given by Mateas is that Bugs Bunny although not a realistic, is however, a believable character. In fact, it has been argued that less realistic or more cartoon like faces have distinct advantages over more realistic ones. Some of these are:

- They have an artistic touch, which makes them more appealing than just seeing a real face (McCloud 1994; Ruttkay, Noot et al. 2000).
- User expectations from less realistic faces are lower than from realistic faces (Microsoft 1999; Ruttkay, Noot et al. 2000).
- Usually less computationally demanding (Ruttkay and Noot 2000; Ruttkay, Noot et al. 2000).
- Exaggerated expressions are easier to achieve on less realistic faces and thus it is easier to convey emotions (Ruttkay and Noot 2000).

Another point worth considering is what we could call communication paradigms. When dealing with anthropomorphic agents we tend to think of 'face-to-face' communication as the only option, however, there are others. One interesting proposition is that of presentation systems where several agents interact with each other on screen to convey information. Elizabeth André (André 2000; André and Rist 2000; André, Rist et al. 2000) is heading research on automatically generating just such performances for two different applications, car sales and soccer commentary. This kind of presentation could prove very successful in the commercial world. Another interesting development is the release of Ananova (Ananova-URL; Hopper 2000) as a mainstream newsreader. Ananova is a computer-generated newsreader. Even though Ananova is no more interactive than a cartoon, it is interesting that the makers of Ananova have decided that the time is right to introduce a computer character in a serious role that was only performed by humans before. It has proven to be a very successful service and it would be of interest to see if similar projects appear in the near future and how people react to them.

2.2.2 Gestures

Gestures are small physical 'activities' that have certain characteristics, which distinguish them from other kinds of activity (such as practical actions, posture adjustments, orientation changes, etc.) (Kendon 1999). Gesture and emotion should not be confused; gestures are voluntary actions, like spoken utterances. It is generally thought that they can expand upon and clarify the ideas and information being transmitted by speech.

There is research being carried out on automatically generating gestures from speech data (Cassell, Pelachaud et al. 1994). There is, however, a lot of controversy still surrounding the way in which gestures are related to speech and activity. Some researchers appear to consider that gestures are simply a kind of symptom from the effort of speaking, others argue that it helps the speaker to speak somehow, yet others believe that they are intrinsically linked to the linguistic choices a speaker makes as he constructs an utterance. Yet another view is that gestures are a separate and distinct mode of expression which can be used together with spoken utterances to complement each other (Kendon 1999). In any case, gestures are an integral part of face to face communication and therefore something which we cannot afford to ignore in our experiments. We will start by concentrating on learning and investigating the interaction between speech and gesture (Cassell, Pelachaud et al. 1994) and how it could be applied to on screen characters to make them more believable agents.

An area currently being researched is the information encoded by gestures and how this information helps the speaker (agent) communicate better (Efron 1972; Rimé 1991; Kendon 1999). Kendon lists the main ways in which gestures appear to be used, these are:

- utterances on their own.
- they may be employed as components of utterances in alternation with speech.
- they may be employed in conjunction with speech.

Each of these will now be discussed briefly:

Gesture alone accompanied by no other utterance: When gestures are employed on their own as an utterance they tend to be of a conventional form that is understood by the whole community. These types of gestures tend to be dependent on culture. For example, a nodding can be used mean yes by many cultures.

Gestures used in alternation with speech: Sometimes speakers can use separate utterances during speech immediately after they finish speaking. Sometimes the speaker might even leave a sentence unfinished and gesture to complete it.

Gestures in conjunction with speech: This is commonly known as 'gesticulation' and is what we will be examining more closely. We are especially interested in the relationship between speech and gesticulation and how it can be modelled to create systems which automatically generate facial and hand gestures (Mulder 1996) for automated conversational agents.

These 'gesticulations' have been subdivided into categories as well, Cassell gives the following classification (Cassell 2002):

- **Emblems,** these are culturally dependant gestures that may differ in interpretation from culture to culture. Cassell gives the following example, the 'V' sign in the US can be made with either with the palm or the back of the hand facing the listener. In

Britain, however, a 'V' sign made with the back of the hand facing the listener is rather rude. Other emblems include 'thumbs up' and ok sign.

- **Propositional Gestures.** These are conscious gestures that in some way describe what the speaker is saying. For instance, if the speaker says "it was this big" and signs the size with his hands, that is a propositional gesture.
- **Spontaneous Gestures,** these are the majority of gestures, those that are unconscious yet serve as gestural vehicles for our communicative intent with other humans. This group of gestures is the most important to study if researcher endeavour to produce believable conversational agents. These types of gestures are further divided into four categories:
- **Iconic** gestures depict some property or event. For instance a gesture to indicate some sort of splitting or separation when the speaker says something like "it branches out".
- **Metaphoric gestures,** like iconic gestures, are also representational. The difference is that they represent a concept that has no physical form. For instance the speaker might gesture as though he is touching something solid whilst saying something like "these ideas".
- **Deictics** are used to locate things in the physical world. It usually involves pointing to objects or people, not just finger pointing, head movements or pointing with an open hand are commonly used as well.
- **Beat gestures** are small baton like movements that do not change in form with the content of the accompanying speech. It servers to evaluate or accentuate the content of speech.

We finally mention gaze. Gaze is sometimes thought of as being different from gesture. However, I believe it might be considered as being the same or at least very closely related to it. We are interested in finding as much as we can about how and when real humans make eye contact during normal conversation so that we can as closely as possible simulate this on our on screen characters. Complex eye contact tracking systems like the Visual Software Agent (Dohi and Ishizuka 1999) begin to touch the realms of robotics and are beyond the scope of our research. We will however be implementing 'human like gaze' into our characters as best as we can with the facilities available.

2.2.3 *Emotions*

It has been recognised both by the AI community and animators that appropriately timed and clearly expressed emotion are a central requirement for believable agents (Bates 1994). Quoting from Thomas and Johnson (Thomas and Johnson 1981):

> *"it has been the portrayal of emotions that has given the Disney characters the illusion of life"*

It can be argued that a character that does not display any emotion cannot be seen by humans as anything more than a machine. The way a character reacts to what is happening around it and displays feelings and desires, makes us care about that character (Bates 1994).

Oz project (Mateas 1997) is concerned with building a small simulated world populated by self animating creatures based on the principles of traditional animation as stated by Thomas and Johnson. When it comes to emotions the Oz philosophy is:

> *"characters exhibit their own emotions and respond to the emotions of others in personality-specific ways"*

This emotional component of the Oz philosophy is key to creating believable agents.

We have seen that emotion is considered to be key to creating believable agents, however, emotion is one of the least well-understood aspects of psychology that we are investigating in this series of experiments. Trying to give a definitive definition of emotion is beyond the scope of this work, however, we place special emphasis on Ekman's basic emotion ideas (Ekman 1993; Ekman 1999) and the Ortony, Clore and Collins (OCC) model of emotion (Ortony, Clore et al. 1988) both of which have been widely used in the anthropomorphic agents field (Koda 1996; Pelachaud, Badler et al. 1996; Velásquez 1997; André, Klesen et al. 1999; Koda and Maes 1999).

Before examining basic emotions in more depth; let us first discuss how Ekman views emotions. Ekman believes there are separate discrete emotions and that all emotions are basic, this is in contrast to a different point of view which treats emotions as being fundamentally the same, differing only in terms of intensity or negativity.

Ekman has found that high agreement on how people interpret emotions across eastern and western cultures and that the human face is key to identifying these emotions.

Ekman (Ekman 1993) has described characteristics, which enable us to distinguish one emotion from another. He has also described characteristics shared by all emotions, which are useful in distinguishing emotions from other affective phenomena, such as moods or emotional trails:

Ekman believes that it was central to the evolution of emotions that the enlighten conspecifics, without choice or consideration, as to what is occurring. For example if you were about to walk into a building and you saw a group of people running out of the building with an expression of fear on their faces you would not think they are just popping out to lunch. He believes that emotions are crucial toward developing and regulating interpersonal relationships. He gives the following examples, which seem to indicate that an emotional component is probably crucial for generating fruitful relationships between users and software agents:

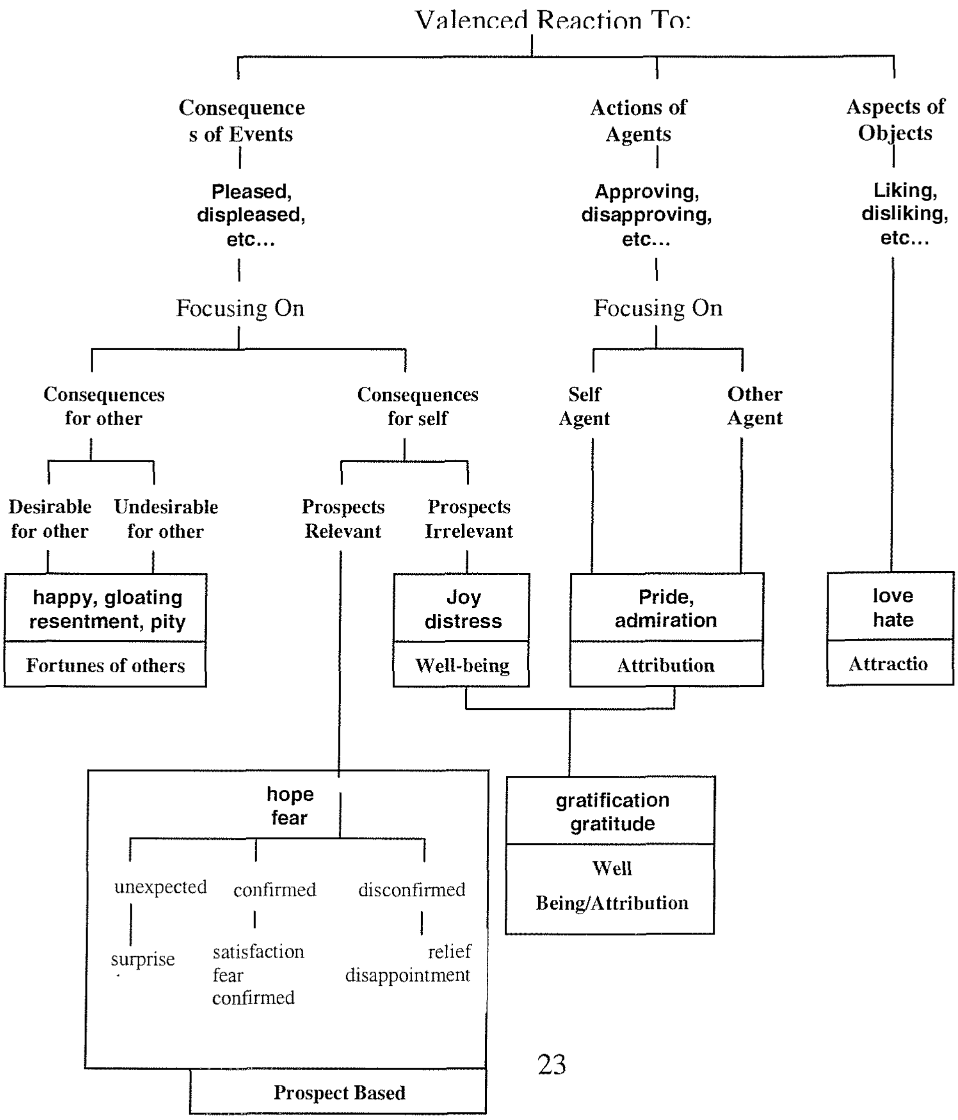


Figure 2.6: OCC model of emotion.

People he has studied with congenital facial paralysis (Mobious syndrome) report that they have great difficulty in developing and maintaining even casual relationships, since they are incapable of displaying facial expressions.

Ross (Ross 1981) has found that stroke patients who can not properly identify the prosody that is associated with speech or who are unable to generate the prosody that accompanies emotion utterances, have severe interpersonal difficulties.

Moods and emotions do not have their own distinctive signals, but Ekman infers these affective phenomena, in part at least, from the fact that they are saturated with the signals of one or another emotion. For example, a high incidence of anger-related signals could indicate an irritable mood or a hostile trait.

Even though it is crucial to the evolution of emotions to inform conspecifics of matters of import, that is not to say that emotions cannot be hidden, masked or faked. However, there is evidence that facial expressions differ in small ways when a smile occurs involuntarily due to an enjoyable experience as opposed to a social or fake smile. Ekman also points out that he suspects if research were carried out on other emotions it would also be possible to distinguish faked emotions and facial expressions from real ones.

The OCC model (See Figure 2.1) is designed to model the structure of emotion; it provides a classification scheme for common emotion labels based on a valence reaction to events and objects in the light of agent Goals, Standards, and Attitudes. And it has been used in the anthropomorphic agents field as a computational model before (André, Klesen et al. 1999).

Koda (Koda 1996) gives the following example: when we compare lottery numbers to the numbers we have selected, this event leads to a pleased emotion if we win or a displeased emotion if we loose. In the OCC model the world consists of events, agents and objects. Valanced reactions to combinations of these factors results in emotions, which can be either positive or negative.

2.2.4 *Creating the onscreen characters*

To create the on screen characters in our work the Microsoft Agent (Microsoft 1999) software was used. Microsoft Agent is a set of programmable software services that support the presentation of software agents as interactive personalities within the Microsoft Windows environment. It has been designed to allow developers to incorporate anthropomorphic conversational interfaces into their software. Early on the possibility of scripting a (Verbot-URL) to create an on-screen character was considered. This, however, proved unsuccessful as the scripting language that the Verbot software uses is very limited and the only way to enable communication between the Verbot and any other software is by using the command line, as the Verbot has been given the ability to load other programs. This is unfortunate as the Verbot has exceptional lip synchronisation and a wide range of facial expressions. As I was impressed by the character design used by the Verbot software that I decided to capture the frames and animate them using the Microsoft Agent software.

As there are no generally accepted guidelines for creating conversational, anthropomorphic interface agents then this is an area we have investigating in this piece of research. Microsoft has produced a set of guidelines for designing effective conversational software agents (Trower 1997):

- **Use good aesthetics**. While it does not take much to elicit a social response, people are accustomed to high quality output.
- **Provide good feedback.** Any well-designed interface requires on providing good, appropriate, and timely feedback. Similarly, a conversational interface, enhanced with recognisable facial features, arose interaction and expectations.
- **Be non-exclusive**. Delegation and speech interface are not always the best way to complete a task. Provide users with different options to interact with the software (Multimodality).
- **Be polite**. Software often breaks simple rules of etiquette. In any social context, humans expect reciprocity of politeness.
- **Use praise**. All of us respond to praise, even when it is not warranted.
- **Be personal**. Adapt to the user's personality type and work style.
- **Create a team player**. We work more effectively with others when we feel we are matched as a team. Treat the user as a team mate.

- **Provide good error handling**. Communications errors are unavoidable, so an effective interface depends on how well you support error recovery.
- **Clarify and limit choices to context**. Speech technology rapidly breaks down when an engine is expected to handle any utterance, yet natural dialogue typically does not follow random paths. Where limitations must be set, they must be clear to the user.
- **Use natural dialogue techniques**. Speech is a natural interface, yet basic models of dialogue are often overlooked.

As mentioned earlier, these are not established guidelines, but they were given consideration when designing the on-screen characters. In this thesis we endeavour to extend these or design a new set of guidelines, backed up with solid research, for interactive character design.

3 Statistical background

3.1 Introduction

This chapter presents a review of the main statistical concepts used in this work (for further details see (Siegel and Castellan 1988)). The concepts are primarily associated with the area of nonparametric statistics for the behavioural sciences and were used in this work to analyse the responses of the test subjects to the different experiments carried out in order to study the user's perception of anthropomorphic agents.

An important part of statistics is the concept of statistical inference. Statistical inference is concerned with two types of problems: estimation of population parameters and tests of hypotheses. It is the latter, tests of hypotheses, which will be our primary concern in this work.

In statistical inference we are concerned with how to draw conclusions about large groups of subjects, or about events, based on the observation pf a few subjects. Statistics provides tools that formalize and standardize our procedures for drawing such conclusions.

The procedures on statistical inference introduce order into any attempt to draw conclusions from the evidence provided by samples. The logic of the procedures dictates some of the conditions under which the evidence must be collected, and statistical tests determined whether, from the evidence we collect, we can have confidence in what we conclude about the larger group from which only a few subjects were sampled.

A common problem for statistical inference is to determine, in terms of a probability, whether observed differences between two samples signify that the populations sampled are themselves really different. Now, even if we collect two groups of scores by taking random samples from the same population, we are likely to find that the scores differ to some extent. Differences may occur simply because of the operation of change. How can we determine in any given case whether the observed differences between two samples are due merely to chance or are caused by other factors? The procedures of statistical inference enable us to determine whether the observed differences are within the range that easily could have occurred by chance or not. Another common problem is to determine whether sample of scores is from some specified population. Still another problem is to decide whether we may legitimately infer that several groups differ among themselves.

There are two main approaches to deal with this type of statistical analysis: one in which it is necessary to make a series of assumptions about the nature of the populations from which the observations or data were drawn. These statistical techniques are called parametric. The other approach makes few assumptions about the population from which the data has been sampled. These distribution free or *non-parametric* techniques result in conclusions, which require fewer qualifications.

The use of one of the non-parametric approaches enables us to infer some conclusions about the event regardless of the shape of the population. When we use any statistical test, we implicitly make certain assumptions about the observations of the event. Here we will describe the main assumptions implied by the tests utilized in this work.

3.2 Statistical tests

In order to analyse how a user perceives an anthropomorphic agent according to its look and behaviour a series of experiments were carried out. In each case a test hypothesis was assumed, which synthesize the results of the experiment. Having stated a specific hypothesis, which seems important, we analyse the data which should enable us to make a decision concerning the hypothesis. Our decision may lead us to retain, revise, or reject the hypothesis.

To reach an objective decision as to whether a particular hypothesis is confirmed by a set of data, we must have an objective procedure for either rejecting or accepting that hypothesis.

This objective procedure should be based on the information or data we obtain in our research and on the risk we are willing to take that our decision concerning the hypothesis may be incorrect.

Usually the objective procedure consists of the following steps:

i. State the null hypothesis (H_0) and its alternative (H_1). Decide what data to collect and under what conditions. Choose a statistical test (with its associated statistical model) for testing H_0.

   In the decision making procedure it is necessary to state the null hypothesis (H_0). The null hypothesis is a hypothesis of "no effect" and is usually formulated for the express purpose of being rejected; that is the negation of the point one is trying to make. If it is rejected, the alternative hypothesis (H_1) is supported. The alternative hypothesis is the operational statement of the experimenter's research hypothesis. The research hypothesis is the prediction derived from the theory under test. When we want to make a decision about differences, we test H_0 against H_1. H_1 constitutes the assertion or hypothesis that is accepted if H_0 is rejected.

ii. From among the several tests which might be used with a given research design, choose that test the model of which most closely approximates the conditions of the research in terms of the assumptions on which the test is based.

For most cases there is an alternative valid statistical test(s) which may be used to reach a decision about a hypothesis. In the field of behavioural sciences non-parametric statistics is most commonly used.

iii. Specify a significance level (α) and a sample size (N). When the null hypothesis and the alternative hypothesis have been stated, and when the appropriate statistical test has been selected, the next step is to specify a level of significance (α) and a sample size (N).

Our procedure is to reject H_0 in favour of H_1 if a statistical test yields a value whose associated probability of occurrence under H_0 is equal to or less than some small probability, usually denoted α. That probability is called the level of significance. It can be seen, then, that α gives the probability of mistakenly or falsely rejecting H_0.

iv. Find the sampling distribution of the statistical test under the assumption that H_0 is true. The sampling distribution is a theoretical distribution. It is that distribution we would have if we took all possible samples of the same size from the same population, drawing each randomly. By knowing the sampling distribution of some statistic, we can make statements about the probability of the occurrence of certain numerical values of a statistic.

v. On the basis of (ii), (iii) and (iv) above, define the region of rejection for the statistical test. The region of rejection consists of a set of small (i.e. equal to α) that the sample we actually observe will yield a value, which is among them. The probability associated with any value in the region of rejection is equal to or less than α. The nature of the rejection is affected by the form of the alternative hypothesis H_1. If H_1 also indicates the predicted direction of the difference, then a one-tailed test is used. If H_1 does not indicate the direction of the predicted difference, then a two-tail test is used. One-tailed and two-tailed tests differ in the location (but not the size) of the region of rejection.

vi. Collect the data. Using the data obtained from the sample(s), compute the value of the test statistic. If that value is in the region of rejection, the decision is to reject H_0; if that value is outside the region of rejection, the decision is that H_0 cannot be rejected at the chosen level of significance. The reasoning behind this decision process is very simple. If the probability associated with the occurrence under the null hypothesis of a particular value in the sampling distribution is very small, we may explain it by deciding that the null hypothesis is false or, second, we may explain it by deciding that a rare and unlikely event has occurred. In the decision process we choose the first of these explanations. Occasionally, of course, the second might be the correct explanation. In fact, the probability associated with the second explanation is given by α.

3.3 Statistical model

In order to choose an appropriate statistical test it is necessary to consider the manner in which the sample of scores or data was drawn, the nature of the population from which the sample was drawn, the particular hypothesis we wish to test, and the kind of measurement or scaling that was employed in the operational definitions of the variables involved, i.e. in the scores.

When we have asserted the nature of the population and the manner of sampling, we have established a statistical model. Associated with every statistical test is a model and a measurement required. The test is valid under certain conditions and a model and the measurement requirement specify those conditions. Often the conditions of a particular statistical model need to be assumed and they are called the assumptions of the test. It is obvious that the fewer or weaker the assumptions that define a particular model, the less we need to qualify the decision reached by the statistical test associated with that model; that is, the fewer of weaker the assumptions, the more general are the conclusions. If those assumptions are valid, tests based on these assumptions are most the likely of all tests to reject H_0 when H_0 is false, in other words, when research data may be analysed appropriately by a parametric test, that test will be more powerful than any other. However, sometimes it is not possible to define at prior these conditions or assumptions and it is necessary to use a more general non-parametric test. The most common conditions considered when dealing with statistical models associated with the normal distribution are:

- The observations must be independent, i.e., the selection of any case from the population for inclusion in the sample must not bias the chances of any other case for inclusion, and the score that is assigned to any case must not bias the score that is assigned to any other case.
- The observations must be drawn from normally distributed populations.
- In the case of analyses concerning two groups, the populations must have the same variance (or, in special cases, they must have a known ratio of variances).
- The variables must have been measured in at least as interval scale, so that it is possible to interpret the results.

These conditions are ordinarily not tested in the course of the performance of a statistical analysis. Rather, they are presumptions, which are accepted, and their truth or falsity determines the accuracy and meaningfulness of the probability statement arrived at by the parametric test. When we have reasons to believe that these conditions are met in the data being analysed, then we should certainly choose a parametric statistical test. When the assumptions constituting the statistical model for a test are, in fact, not met, then the test may not be valid, i.e., a test statistic may fall in the rejection region with a probability greater than α. In those cases it is worthy to consider the alternative of non-parametric statistics.

3.4 Parametric and non-parametric statistics

A parametric statistical test specifies certain conditions about the distribution of responses in the population from which the research sample was drawn. Since these conditions are not ordinarily tested, they are assumed to hold. The meaningfulness of these results of a parametric test depends on the validity of these assumptions.

On the other hand, a non-parametric statistical test is based on a model that specifies only very general conditions and none regarding the specific form of the distribution from which the sample was drawn. Certain assumptions are associated with most non-parametric statistical tests, namely, that the observations are independent, perhaps that the variable under study has underlying continuity, but these assumptions are fewer and weaker than those associated with parametric tests. Moreover, unlike parametric tests, there are non-parametric tests that may be applied appropriately to data measured in an ordinal scale, and others to data in a nominal or categorical scale.

Because behavioural scientists rarely have data satisfying the assumptions of the parametric test, which includes achieving the sort of measurement permitting the meaningful interpretation of parametric tests, non-parametric statistical tests play a prominent role in research in behavioural sciences.

The non-parametric tests used in this work are:

- Chi-square goodness of fit method.
- Friedman two-way analysis.
- Wilcox-Mann-Whitney.

The main advantages of using non-parametric tests are:

- If the sample size is very small, there may be no alternative to using a non-parametric statistical test unless the nature of the population distribution is known exactly.
- Non-parametric tests typically make fewer assumptions about the data and may be more relevant to a particular situation.
- Non-parametric statistical tests are available to analyse data, which are inherently in ranks, as well as data whose seemingly numerical scores have the strength of ranks. If data are inherently in ranks, or even if they can be categorized only as plus or minus (e.g. more or less, better or worse), they can be treated by non-parametric methods, whereas they cannot be treated by parametric methods unless precarious and, perhaps, unrealistic assumptions are made about the underlying distributions.
- Non-parametric methods are available to treat data that are simply classificatory or categorical, i.e., are measured in a nominal scale. No parametric technique applies to such data.
- There are suitable non-parametric statistical tests for treating samples made up of observations from several different populations. Parametric tests often cannot handle such data without requiring us to make seemingly unrealistic assumptions or requiring cumbersome computations.
- Non-parametric statistical tests typically are much easier to learn and to apply than are parametric tests. In addition, their interpretation often is more direct than the interpretation of parametric tests.

However, non-parametric tests can have the following disadvantages:

- If all the assumptions of a parametric statistical model are met in the data and the research hypothesis could be tested with a parametric statistical test, then non-parametric statistical tests are in these cases wasteful.
- Non-parametric tests are not systematic, whereas parametric statistical tests have been systematized, and different tests are simply variations on a central theme.
- Tables necessary to implement non-parametric tests are scattered widely and appear in different formats.

3.4.1 Chi-square Goodness-of-Fit Test

Frequently research is undertaken in which the researcher is interested in the number of subjects that fall into a certain category. For example, in the two-way interaction experiment in Chapter 4 the users were asked to choose a favourite character between three different characters with different level of anthropomorphic abstraction, from very abstract to a realistic yet not human character. Amongst the users one of the characters is picked out as being the favourite character more frequently than the others. The degree of interest of a character may be categorized according to the frequency that they were chosen as favourite by the users, the hypothesis being that these characters will differ in frequency in a prescribed way. The experiment was planned to test the hypothesis that the degree of interest of the characters differs in frequency.

The chi-square test is suitable for analysing data like these. The number of categories may be two or more. The technique is of the goodness-of-fit type in that it may be used to test whether or not a significant difference exists between an observed number of objects or responses falling in each category and an expected number based upon the null hypothesis. That is, the chi-square test assesses the degree of correspondence between the observed and expected observations in each category.

To compare an observed with an expected group of frequencies, we must be able to state what frequency would be expected. The hypothesis H_0 (the null hypothesis) states the proportion of objects falling in each of the categories in the presumed population, i.e., in our example each of the characters are considered to have the same frequency 33.33%. It is important to remember that the null hypothesis is formulated for the express purpose of being rejected. Therefore, we are looking for the possibility of popularity being different between the characters, i.e., our alternative hypothesis (H_1).

The chi-square technique gives the probability that the observed frequencies could have been sampled from a population with the given expected values. The null hypothesis H_0 may be tested using the following statistic:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Equation 1

Where O_i = the observed number of cases in the ith category.

E_i = the expected number of cases in the ith category when H_0 is true.

k = the number of categories.

Thus Equation 1 directs one to sum over k categories the squared differences between each observed and expected frequency divided by the corresponding expected frequency.

If the agreement between the observed and expected frequencies is close, the differences (O_i - E_i) will be small and, consequently, X^2 will be small. However, if the divergence is large, the value of X^2 as computed from Equation 1 also will be large. Roughly speaking, the larger the value of X^2, the less likely it is that the observed frequencies came from the population on which the hypothesis H_0 and the expected frequencies are based.

Although Equation 1 is useful for understanding the X^2 statistic, it is often cumbersome to compute because of the number of subtractions involved. After some manipulation, a somewhat more convenient computing formula can be found:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$
$$= \sum_{i=1}^{k} \frac{O_i^2}{E_i} - N$$

Equation 2

where N is the total number of observations.

It can be shown that the sampling distribution of X^2 under H_0, as computed from Equation 1, follows the chi-square distribution with degrees of freedom $df = k - 1$.

There are a number of different sampling distributions for chi-square, one for each value of *df,* the degrees of freedom. The size of *df* reflects the number of "observations" which are free to vary after certain restrictions have been placed on the data. For example in our previous case $df = 2$, in general, for the one-sample goodness-of-fit test, when H_0 fully specifies the expected frequency, $df = k - 1$, when k is the number of categories in the classification; *3* characters in our case.

To use chi-square in testing a hypothesis in the one-sample goodness-of-fit situation, cast each observation into one of k cells. The total number of such observations should be N, the number of cases in the sample. That is, each observation must be independent of every other. For each of the k cells, the expected frequency must also be entered. If H_0 is that there is an equal proportion of cases in each category in the population, then $E_i = N/k$. With the various values of E_i and O_i known, one may compute the value of X^2 by the application of Equation 1. The significance of this obtained value of X^2 may be determined by reference to the chi-square distribution usually given in tables. If the probability associated with the occurrence under H_0 of the obtained X^2 for $df = k-1$ is equal to or less than the previously determined value of α, the H_0 may be rejected. If not, H_0 cannot be rejected.

The result of the above experiment as well as other experiments that use the chi-square test is given in subsequent chapters. The numerical analysis in these cases for this type of test were obtained using SPSS for Windows (Statistical Package for Social Sciences).

3.4.2 Wilcoxon-Mann-Whitney test

When at least ordinal measurements has been achieved for the variables studied, the Wilcoxon-Mann-Whitney test may be used to test whether two groups have been drawn from the same population. This is one of the most powerful of the non-parametric tests. Suppose we have samples from two populations, X and Y. The null hypothesis is that X and Y have the same distribution. The alternative hypothesis H_1 against which we test H_0 is the X is stochastically larger that Y – a directional hypothesis. We may accept H_1 if the probability that the score from X is larger than a score from Y is greater than one-half. That is, if X is one observation from population X and Y is an observation from the population Y, then H_1 is that $P[X>Y]>1/2$. If the evidence supports H_1, this implies that the "bulk" of the elements of population X are larger than the bulk of the elements of population Y. Using this approach, the null hypothesis is H_0: $P[X>Y]=1/2$.

Of course, our hypothesis might instead be that Y is stochastically larger that X. In that case, the alternative hypothesis H_1 would be that $P[X>Y]<1/2$. Confirmation of this assertion would imply that the bulk of Y is larger that the bulk of X.

For a two-tailed test, i.e., for a prediction of differences which does not state the direction of the differences, H_1 would be that $P[X>Y] \neq 1/2$.

Another way of stating the alternative hypothesis is that the median of X is greater than the median of Y, that is, H_1: $\theta_x > \theta_y$. In a similar fashion, the other hypothesis may be stated in terms of medians.

Therefore this type of test can be used to determine the number of times a score from one of the samples is ranked higher than a score from the other sample. If the two sets of scores are similar, then the number of times this happens should be similar for the two samples.

For example in Chapter 6, we compare how the users perception of a character is affected by the characters display of emotions using two different sets of user samples in which one of the sets the characters exhibits emotions and in the other it does not.

In this case the null hypothesis, H_0 is given by the condition that the users' perception is unaffected by the display of emotion, while the alternative hypothesis, H_1, states that the character's display of emotions does indeed affect users' perception.

To apply the Wilcoxon test, we first combine the observations or scores from both groups and rank them in order of increasing size, where m observations or cases are from the group x and n from the group y. In this ranking, algebraic size is considered, i.e., the lowest ranks are assigned to the largest negative values, if any.

To find W_x, we first rank the score's identity as either an X or Y score. In a similar way the value of W_y is computed. The sum of the ranks for the two groups should be equal to the sum of the ranks for the combined group. That is,

$$W_x + W_y = \frac{N(N+1)}{2}$$

Equation 3

If H_0 is true we would expect the average ranks in each of the two groups to be about equal. If the sum of the ranks for one group is very large (or very small), then we may have reason to suspect that the samples were not drawn from the same population. The sampling distribution of W_x when H_0 is true known, and with this knowledge we can determine the probability associated with the occurrence under H_0 of any W_x as extreme as the observed value.

The Wilcoxon test assumes that the scores are sampled from a distribution that is continuous. With very precise measurement of a continuous variable, the probability of a tie is zero. However, with the relative crude measures that are typically employed, ties may occur. We assume that the two (or more) observations which result in tied scores are really different, but that this difference is simply too refined or minute to be detected by our measurements.

When tied scores occur, we give each of the tied observations the average of the ranks they would have had if no ties had occurred.

If the ties occur between two or more observations in the same group, the value of W_x is not affected. But if the ties occur between two or more observations involving both groups, the value of W_x (and W_y) is affected. Although the effect is usually negligible, a correction for ties is available and should be used whenever we employ the large-sample approximation to the sampling distribution of W_x. The effect of tied ranks it to change the variability if the set of ranks. Thus the correction for ties must be applied ti the variance of the sampling distribution of W_x. Corrected for ties, the variance becomes:

$$\sigma_{W_x}^2 = \frac{mn}{N(N-1)}\left(\frac{N^3 - N}{12} - \sum_{j=1}^{g}\frac{t_j^3 - t_j}{12}\right)$$

Equation 4

where $N = m + n$, g is the number of groupings of different tied ranks, and t_j is the number of tied ranks in the jth grouping.

When we do not correct for ties our test is "conservative" in that the associated probability will be slightly inflated compared to that for the corrected. That is, the value of the probability associated with the observed data when H_0 is true will be slightly larger than that one that would be found were the correction employed. In general it is recommended that one should correct for ties only if the proportion of ties is quite large, if some of the t's are large, or if the probability obtained without the correction is very close to the set value of α.

The steps in the use of the Wilcoxon test are:

- Determine the value of m and n. The number of cases in smaller group (dented X) is m; the number of cases in the larger group (denoted Y) is n.
- Rank together the scores for both groups, assigning the rank of 1 to the score that is algebraically the lowest. The ranks will range from 1 to $m+n=N$. Assign tied observations the average of the tied ranks.
- Determine the value of W_x by summing the ranks in group X.
- Determine the associated probability of W_x using the corresponding distribution. For small values of m and n this value can be obtained from tables and for large values it is possible to use the normal approximation.
- If the observed value of W_x has an associated probability equal to or less than α, reject H_0 in favour of H_1.

3.4.3 Friedman Two-Way Analysis of Variance by Ranks

When the data from k matched samples are in at least an ordinal scale, the Friedman two-way analysis of variance by ranks is used to test the null hypothesis that the k samples have been drawn from the same population.

Since the k samples are matched, the number of cases N is the same in each of the samples. That matching may be achieved by studying the same group of subjects under each k conditions.

The Friedman two-way analysis of variance by ranks tests the null hypothesis that the k repeated measures or matched groups come from the same population or populations with the same median. To specify the null hypothesis more explicitly, let θ_j be the population median in the jth condition or group. Then we may write the null hypothesis that the medians are the same as H_0: $\theta_1 = \theta_2 = \ldots = \theta_k$. The alternative hypothesis is then H_1: $\theta_i \quad \theta_j$ for at least two conditions or groups i and j. That is, if the alternative hypothesis is true, at least one pair of conditions has different medians. Under the null hypothesis, the test assumes that the variables have the same underlying continuous distribution; thus it requires at east ordinal measurements of that variable.

In Chapter 4 a series experiments were performed designed to determine how levels of anthropomorphism and interaction of three different characters, $k=3$, affects user's perception of anthropomorphic interface agents. The attributes examined were friendliness, intelligence, pleasantness, how interest the characters are and helpfulness. These were examined under three different experimental conditions; No interaction, one-way interaction and two-way interaction between the characters and the users. For each of these experiments the null hypothesis was that the difference in the level of anthropomorphism had no differential effect on the user's perception of each of the character's attributes. The alternative hypothesis states that the difference in level of anthropomorphism has differential effect.

The Friedman test determines whether the ranks totals (denoted R_j) for each condition or variable differ significantly from the values that would be expected by chance. To do this test, we compute the values of the statistic, which we shall denote as F_r.

$$F_r = \left[\frac{12}{Nk(k+1)} \sum_{j=1}^{k} R_j^2 \right] - 3N(k+1)$$

Equation 5

where N = number of rows (subjects)

k = number of columns (variables or conditions)

R_j = sum of ranks in the jth column (i.e. the sum of ranks for the jth variable)

And $\sum_{j=1}^{k}$ directs us to sum the squares of the sums of ranks over all conditions.

Probabilities associated with various values of F_r when H_0 is true have been tabulated for various sample sizes and various numbers of variables. If one observed value of F_r is larger than the predicted value of F_r from the distribution at the chosen significance level, then H_0 may be rejected in favour of H_1.

When the number of rows and/or columns is large it can be shown that the statistic F_r is distributed approximately as χ^2 with $df=k-1$.

When there are ties among the ranks for any given group (or row) the statistic F_r must be corrected to account for changes in the sampling distribution. Equation 6 gives the value of F_r, which is appropriate when ties occur. Although Equation 6 can be used in general, that is, when there are no ties as well as when there are ties, the computation is somewhat more tedious.

$$F_r = \frac{12\sum_{j=1}^{k} R_j^2 - 3N^2k(k+1)^2}{Nk(k+1) + \frac{\left(Nk - \sum_{i=1}^{N}\sum_{j=1}^{g_i} t_{i.j}^3\right)}{(k-1)}}$$

Equation 6

where g_i is the number of sets of tied ranks in the ith group and $t_{i.j}$ is the size of jth set of tied ranks in the ith group. As is the case with other corrections for tied data, the effect of tied ranks is to crease the size of the Friedman statistic F_r.

When the obtained value of F_r is significant it indicates that at least one of the conditions differs from at least one other condition. It does not tell the researcher which one, nor does it tell the researcher how many of the groups are different from each other. That is, when the obtained value of F_r is significant we would like to test the hypothesis H_0: $\theta_u = \theta_v$ against the hypothesis H_1: $\theta_u \neq \theta_v$ for some conditions u and v. There is a simple procedure for determining which condition or conditions differ. First the differences $|R_u - R_v|$ for all pairs of conditions or groups most be determined. When the sample size is large, these differences are approximately normally distributed. However, since there are a large number of differences and because the differences are not independent, the comparison procedure must be adjusted appropriately. The hypothesis of no difference between the k condition or matched groups was tested and rejected at the α significance level. Then the significance of individual pairs of differences can be tested by using the following inequality. That is, if

$$|R_u - R_v| \geq z_{\alpha/k(k-1)}\sqrt{\frac{Nk(k+1)}{6}}$$

Equation 7

or if the data are expressed in terms of average ranks within each condition, and if

$$|\overline{R}_u - \overline{R}_v| \geq z_{\alpha/k(k-1)}\sqrt{\frac{k(k+1)}{6N}}$$

Equation 8

then we may reject the hypothesis H_0: $\theta_u = \theta_v$ and conclude that H_1: $\theta_u \neq \theta_v$. Thus, if the difference between the rank sums (or average ranks) exceeds the corresponding critical value given in Equation 7 or Equation 8, then we may conclude that the two conditions are different. The value of $z_{\alpha/k(k-1)}$ is the abscissa value from the unit normal distribution above which lie $\alpha/k(k-1)$ percent of he distribution.

The steps in the use of the Friedman two-way analysis of variance by ranks are:

- Cast the scores in a two-way table having N rows (subjects) and k columns (conditions or variables)
- Rank the data in each row from 1 to k.
- Determine the sum of the ranks in each column (R_j).
- Compute the value of F_r with Equation 5 if there are no ties or Equation 6 if there are tied observations in any row.
- Determine the probability of occurrence when H_0 is true of an observed value of F_r.
- If the probability yielded in the previous step is equal to or less that α, then reject H_0.
- If H_0 is rejected, use multiple comparisons (Equation 7 and Equation 8) to determine which differences among conditions are significant.

As in the previous two cases the analysis of the Friedman test was carried out using SPSS.

3.5 Statistical Package for the Social Sciences (SPSS)

To understand the results of any experiment it is necessary to carry out statistical analysis of the resulting data. These calculations are relatively easily and quickly carried out by hand when the data consists of a small set of low volume numbers and when the statistics are comparatively straightforward. However, with larger sets of numbers, the manual calculations are time consuming and may result in elementary mistakes being made. This becomes serious when we need to employ more complicated statistics. The advent of computers has led to the development of various computer programs for calculating statistics. These programs don't tend to be difficult to use and carry out calculations very quickly. With the ready accessibility of these packages, calculating statistics with the aid of a computer is generally efficient.

The different statistical analyses carried out in this work were performed with SPSS (Statistical Package for the Social Sciences). The development of SPSS began in 1965 at Stanford University and continued in 1970 at the National Opinion Research

Centre at Chicago University. It became a commercial organisation based in Chicago in 1975. SPSS is readily available on both mainframes and personal computers. As the software is continuously being revised and the and the revisions take time to be implemented in different institutions, various versions of it may be accessible at any one time and these versions may be periodically replaced with new ones when these becomes available.

The first version for personal computers was developed for the Apple Macintosh and latter for Microsoft Windows. The earlier PC version was Release 4 for Windows, we used Release 11 for Windows throughout our research.

For more information on SPSS, see (Cramer 1998).

4 Levels of anthropomorphic abstraction and interaction

This chapter reports on experiments investigating how levels of anthropomorphism and interaction affect users' assumptions when interacting with anthropomorphic interface agents. User trials were undertaken under three different experimental conditions: no interaction, one-way and two-way interaction.

It was demonstrated that the simpler experimental conditions replicated those of the more complex conditions, enabling us to create simpler and quicker experiments for future research.

A number of key patterns in the assumptions made by users as they interacted with changing levels of anthropomorphic abstraction of an interface agent are identified.

The users described the more realistic looking characters as being "scary", although, these realistic characters are seen as being more intelligent. While the users scored the abstract characters as being more friendly and pleasant, they also described them as being less interesting and boring.

The perceived value of each of the characters varied with the level of interaction. From these results we concluded it is better to have no interaction with the characters rather than only partial interaction. The character favoured by the users was one of moderate abstraction.

4.1 Introduction

The aim of this section is to investigate users' reactions and assumptions when interacting with anthropomorphic agents. How the level of anthropomorphism exhibited by the character and the level of interaction affects these assumptions is considered. Characters of different levels of anthropomorphic abstraction, from a very abstract character to a realistic looking character are compared. As more software is released for general use with anthropomorphic interfaces there seems to be no consensus of what the characters should look like and what look is more suited for different applications. Some software and research opts for realistic looking characters

(Haptek-URL), others opt for cartoon characters (Microsoft 1999) while others opt for floating heads (Dohi and Ishizuka 1997), (Takama, Dohi et al. 1998), (Koda and Maes 1999).

On anthropomorphic abstraction Thórisson (Thórisson and Cassell 1996) argues that realistic computer animated faces look abnormal and repulsive. Similarly McCloud (McCloud 1994) proposes that cartoon characters are more likely to be accepted by viewers, because, when viewers see a cartoon character they see a reflection of themselves. i.e. cartoons are like an empty shell that enables us to not just watch the cartoon, but to "become it".

The way the level of interaction between the user and the agent affects these assumptions was also investigated. The results of this investigation are of particular interest as this will assist in designing future experiments in this area. To realize this three sets of experiments with different levels of interaction between the user and the animated character were carried out:

- A set of four qualification tests (friendliness, intelligence, pleasantness and interestingness) was carried out when no interaction was allowed between the users and characters. Test subjects were just asked to rate the characters with different levels of anthropomorphic abstraction on these four characteristics only by judging their appearance from a still picture.
- The above set of four characteristics was tested for, plus one more characteristic related to how helpful the users found the characters to be was carried out by allowing one-way interaction from character to user. This experiment consisted of a lecture using PowerPoint presented by animated characters exhibiting different levels of anthropomorphic abstraction. Users were then asked to rate these characters' performance. It is important to point out that the helpfulness characteristic was not considered on the previous experiment since no communication was allowed between the user and characters, making this question redundant.
- Finally the above set of five tests was performed allowing two-way interaction between the user and the characters of different levels of anthropomorphic abstraction; from user to character and vice versa. For the user/agent interaction we developed a natural language web-searching interface.

All interaction during the two-way-interaction experiment was recorded in order to ascertain changes in work strategy owing to the different levels of anthropomorphic abstraction of the characters. A test is carried out to determine how politely users act toward the agent, which is used for establishing to what extent users behave toward the agent as they would toward a human.

4.2 Experimental procedure

This section describes the three experiments carried out in chronological order.

4.2.1 The on-screen characters

Three different looking characters were utilized for the first experiment (Figure 4.1), albeit, with exactly the same capabilities. The users were not aware of this, they were simply asked to judge the capabilities of the different agents. See Appendix 1 for frames of animation used for these characters in these experiments.

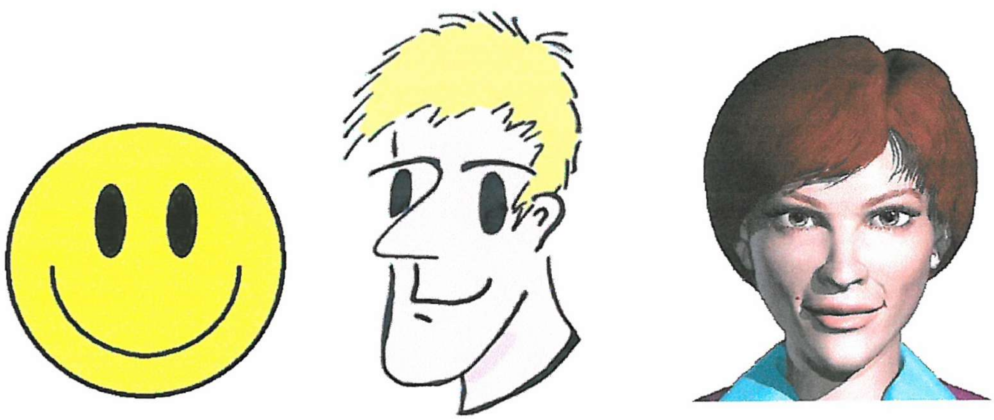


Figure 4.1: Characters used for the first experiment.

After negative user feedback on the first two sets of experiments concerning the female character, she was described as being *"scary"* by a considerable number of test subjects, we decided to change this character for another. This time we used Haptek's Virtual Friend character Erin, which is a smoothly animated 3D character (Figure 4.2)

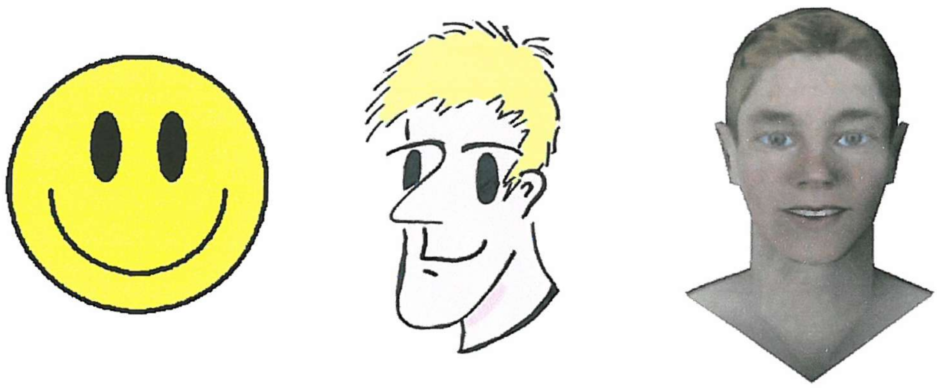


Figure 4.2: Characters used for the third experiment.

4.2.2 Pilot Experiment:

To test the experimental procedure and get user feedback concerning the usability of the software created for the first experiment we carried out a pilot experiment with three test subjects in order to highlight any possible problems with the interface, procedure, experimental setting or instructions provided. This step was considered critical to the success of experiments, as we wanted to minimize the likelihood of setbacks or usability issues that could bias experimental results as much as possible.

For this pilot and the first experiment, code was implemented that works as an interface for a web based search engine (see Figure 4.3) by using Sherlock plug-ins to define the output of search engines so it can be parsed. In this experiment two-way communication was allowed between the user and the character. The ability of this system to target different sources can be exploited to give different capabilities to the agent. For example, if the user asks the agent to find a photograph the agent would only target archives that contain pictures. We restricted the software to a few simple functions: 'find website', 'find picture' and 'find links'. Originally we intended to target several search engines at the same time and then analyse results and rank them in order of relevance, however, this slows down the search process and since Google™ (Brin and Page 1998) produces such good results we decided to only target Google™. The find website function also targets Google™ but selects the most relevant website. For the find picture function, the software targets the Altavista™ image index. However, at the time we found that it didn't produce very good matches so we deactivated this feature for the first experiment. Having these clear functions will enabled the author to instruct users to carry out specific tasks without telling them exactly what to do. For instance, you can instruct the test subject to find a book about a subject that interests them or to find a picture of a pop star. For the first set of tests we are using the Google™ and Netscape™ search engines as our sole sources of links. The natural language parser can handle three types of inputs:

- Commands to find a set of links
- Commands to find one page
- Keywords

The agent parses the first two types of commands and it forwards the search keys to Google™ and waits for the results. A command that cannot be parsed is sent to Netscape Search™, which accepts natural language queries. The agent can also parse commands for finding pictures, however, as stated earlier we deactivated this feature for the first experiment.

To facilitate possible future Wizard of Oz type experiments the test software has been implemented in such a way that it can be remotely controlled by setting up a server that does the language parsing. This server receives the natural language queries typed by the user from the program controlling the character, and then sends it back a command and parameters that lets it know what to do.

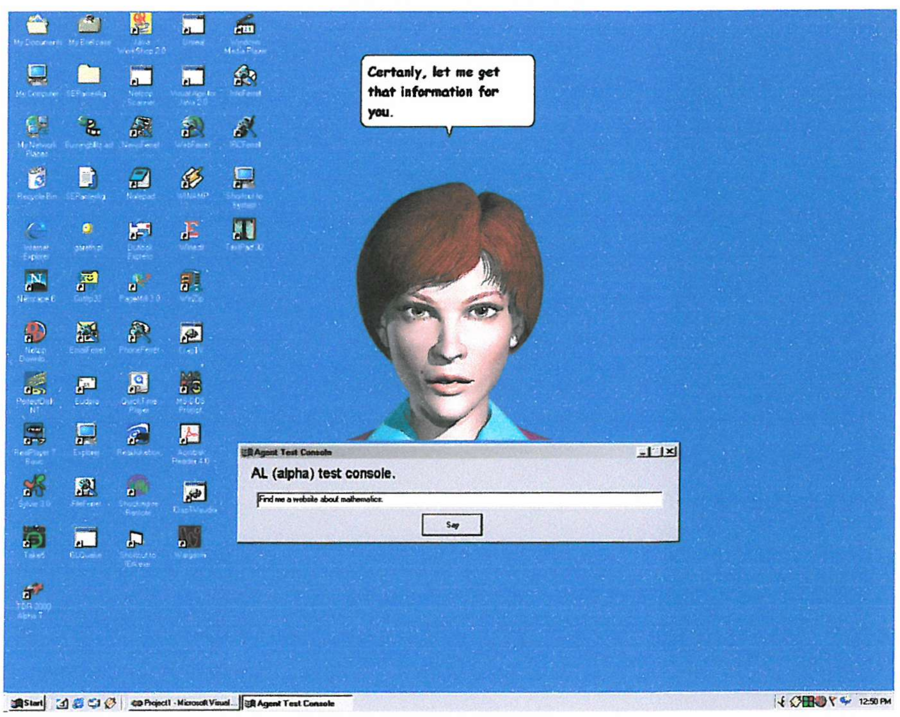


Figure 4.3: Screen shot of our onscreen conversational character running.

A Wizard of Oz experiment is an evaluation method in which the user of the system is made to believe that he or she is interacting with a fully implemented system even though the some or all of the functionality of the system is controlled by a concealed human, i.e.: the wizard.

4.2.2.1 Lessons learned from pilot experiment

Minor complaints about the software were addressed, users wanted to press enter instead of having to use the mouse all the time to input their queries. It was realized that written instructions were needed in order to make sure the experimenter was consistent when instructing the users prior to the experiment.

It was initially intended to support speech recognition for this experiment, however, it was finally decided not to support speech recognition because it was thought that having an extra layer on the interface between the user and the agent, might or might not work reliably. This would most probably compromise our results by making the user feel uncomfortable with the interface. It is important to remember that the purpose of these experiments is to examine the user's reactions to different styles of interactive characters and stay focused on this when designing the experiment. Therefore, speech recognition really is not an integral part of this first trial.

4.2.3 Two-way interaction experiment

In order to observe if some of the features used were not necessary for simpler experiments, we decided to perform the most complex of the experiments first (i.e. the two-way interaction experiment), see Table 4.1 for user profile.

The software used for this experiment is the same as was used for the pilot experiment; however, we addressed the minor complaint concerning the input button. As mentioned earlier this software uses Sherlock plug-ins to define the output and the input expected by different search engines. This ability was exploited to give different capabilities to the software. For example, if the user (See Table 4.1 for user profile) asks the character to find a photograph the software only targets archives that contain pictures. The software was restricted to a few simple functions: find website, find picture and find links. The find website function also targets Google™ but selects the first match which should be the site which Google™ rated most as being relevant. For the find picture function, the software targets the Altavista™ image index. However, during a pilot experiment we found that it didn't produce very good matches so we deactivated this feature. Having these functions enabled us to instruct users to carry out open-ended tasks (see Table 4.2).

Population information	**The two-way interaction experiment was advertised locally at the Department of Electronics and Computer Science at the University of Southampton.**
Number of Users	11
Gender distribution	8 male and 3 female
Age distribution	Age group 18-30

Table 4.1: User profile for the first experiment: two-way interaction.

The natural language parser can handle three types of inputs:

- Commands to find a set of links
- Commands to find one page
- Keywords

The software parses the first two types of command, extracting search keys and forwards these to Google™ and waits for the results. A command that cannot be parsed by our software is sent to Netscape Search, which also accepts natural language queries.

Search for and browse sites on your academic/work interests.
Search for and browse sites on your hobbies.
Search for and browse sites on a music group(s) of your choice.
Search for and browse sites on your favourite sport(s).
Search for information on a company or product of your choice.
Search and browse sites on your favourite TV program(s).

Table 4.2: List of tasks for two-way interaction experiment.

	Attribute	Description
User-friendly interaction.	Friendly	A character for this application should be of favourable disposition and well wishing. Both of which are characteristics, which are encompassed by the word friendly.
	Pleasant	Being pleasant is defined as having pleasing manners, demeanour, and aspect; agreeable, cheerful, good-humoured.
	Interesting	The author wishes to provide the user with an interesting character to interact with and look at. This is to maintain user involvement in the interaction
Fulfilling the user's needs and expectations	Intelligence	Is an important trait that has been investigated before (Koda 1996); it has been thought that matching the anthropomorphic character's outward intelligence to its capabilities is crucial if users are not to be frustrated when the software's capabilities do not meet their expectations.
	Helpful	The author believes that it is vital that an anthropomorphic assistant appear helpful and eager to help.

Table 4.3: Attributes to be examined in the experiments.

To ensure that the user's opinion was not influenced by the tasks they decided to carry out, their actions were restrained by the guidelines given to them.

To ensure that the tests were fair, it was decided to give test subjects a set of searches that needed to be done by querying the software. In addition, to ensure that the results were not influenced by the order in which these tasks were undertaken, the order was randomised using a random table (Fowler et al., 1998).

A list of six open-ended tasks was produced, which were divided into three sets of two tasks (see Table 4.2). These were printed out and handed to the users as required, when they finished one set of tasks another was given. When the user finished a set of tasks the next would be given to them and the next character they had to interact with would be started. To further reduce the chances of the experimental procedure affecting the results we randomised the order in which the characters were presented to the users, once again using the random table. In this experiment the characters used had a neutral expression all the time (i.e. they did not show any emotion).

The users' opinions were recorded using a questionnaire (see Appendix 2); this was completed once they had carried out all the tasks as instructed. Here they were asked to rate each character (see Figure 4.1) using a seven-point scale for each of the attributes, 1 being the lowest (i.e. un-Helpful) and 7 the highest (i.e. Helpful).

In this experiment users were asked to rate a set of attributes. It was decided to identify traits considered central to the application of a simple anthropomorphic web-searching assistant. We concentrated on two goals that were thought to be essential towards forming a fruitful relationship between the user and character and were dictated primarily by its appearance (see Table 4.3):

- User friendly interaction: It was thought that interaction with the character should be an enjoyable experience, to maximize these enjoyment it was thought the character should seem to be friendly and pleasant.
- Fulfilling the user's needs and expectations of the software.

These attributes can be seen as soft-goals (Mylopoulos, Chung et al. 2001). Not all goals can be defined clearly nor can they have clear-cut criteria to determine whether they have been satisfied. Table 4.1 in other words represents a partial soft goal hierarchy designed to satisfy user expectations.

The experiment took place in a laboratory and the users were fully instructed (see Appendix 3 for instructions given to users) on how to communicate with the characters and how to interpret the results.

4.2.3.1 Lessons learned from the two-way interaction experiment

Some users commented on how the speech bubbles caused them to concentrate on the text instead of looking and listening to the character. It was decided that for the subsequent experiment speech bubbles would be omitted, as in these more expressions would be displayed by the characters, which were important for the test subjects to see. The practicalities of having the user stop to carry on with the test were discussed and it was decided that it would be better to design the software so that there is no need for the users to stop and request further tasks. For example, the software could provide the user with the tasks to be carried out and it could change the character used for the interface as the user completes each task.

4.2.4 No Interaction Experiment

The second experiment consisted of a web-based form where users (see Table 4.4 for user profile) were asked to assess the characters just by looking at still pictures i.e. no interaction.

The helpfulness attribute was omitted as it was thought that it wasn't fitting to judge how helpful a person was from a still picture.

Population information	The no interaction experiment was advertised across the University of Southampton.
Number of Users	31

Table 4.4: User profile for the no-interaction experiment.

The complete online form used for data collection in this experiment can be seen in Appendix 4.

4.2.5 *One-way interaction experiment*

The third experiment consisted of a lecture presented to undergraduate students (see Table 4.5) by three different characters (see Figure 4.2). As described previously in 4.2.1 it was decided to change the realistic character for this experiment, as the feedback from users for the previous two experiments was negative. The female character was described as "scary" by a considerable number of users, both in written and verbal comments after the experiment. To replace the character we used Haptek's Virtual Friend Eric character from Haptek's Virtual Friend software. The main part of the presentation was divided up into three parts. Part one took 12 minutes, part two took 7 minutes and part three took 5 minutes. The first part was presented by character A in Figure 4.2, part two by character B and part three by character C. It was decided that each part should take less time than the previous one to prevent the students from getting bored and irritated with the presentation. The attributes we examined for each of the characters are shown in Table 3.

For this experiment we had two expressions: neutral and smile. The characters would speak with a neutral expression and smile from time to time. Character C is slightly different from characters A and B, Haptek's character is rendered in 3D and has smoother animation when compared to the other two.

Population information	In the one-way interaction experiment users were obtained from a lecture on Human Computer Interaction at the Department of Electronics and Computer Science at the University of Southampton.
Number of Users	79
Gender distribution	77 male and 2 female
Age distribution	Between 18-25

Table 4.5: User profile for the third experiment: one-way interaction.

The attributes we examined for the characters were:

- Friendliness
- Intelligence
- Pleasantness

- How interesting they are
- Helpfulness

Characters A and B have exactly the same voice and character C has a similarly sounding voice. Haptek's technology at the time did not allow for the use of different text to speech engines whereas Microsoft's Agent technology does, so we chose the voice we thought sounded the most similar to Haptek's choice of voice, Microsoft's text-to-speech engine Mike.

As all the characters did not have exactly the same voice an experiment was carried out to find out how much the voice influences the test subject's opinions. Three voices were chosen; Microsoft text-to-speech engine Mike, Digalo's (developed by Elan) Gordon voice that has a British accent and AT&T Bell Labs adult male 1 (which has an American accent). Three small summaries of the lecture were read by character B, the cartoon character, (figure 1, 2) but with a different voice for each summary. We asked the students to evaluate the cartoon character when the different voices were used to assess how their appreciation of the character was influenced by the voice.

The use of a voice actor was considered as users reported that text-to-speech (TTS) sounds unnatural and is unpleasant to listen to (Ralston, Pisoni et al. 1995). However, we were also interested in determining if the TTS voices were suitable for a teaching application, using TTS instead of recorded voice actors facilitates the quick production of automated lecture material. There is strong evidence that although TTS still does not match the clarity and prosody of normal human speech (Olive 1997), however, word intelligibility scores for the best TTS systems approaches that of human speech (Kamm, Walker et al. 1997) and can accurately manifest personality as well (Nass and Lee 2000).

4.2.5.1 Lessons learned from the one-way interaction experiment

The nature of this experiment meant that we had to accept the fact we were more likely to create small biases when compared to the previous experiments. For instance there is nothing we could do about the order in which the characters were presented, the characters presented different sections of the lecture and as the novelty effect wears out the test subjects could get bored towards the end of the lecture.

There were two mistakes on the printed questionnaire forms. The age group 46-50 was missing, and a spelling mistake on the third page was evident ("your" instead of "you"). Some users remarked that at first they were confused as the characters on the forms were in a different order to which they were presented in the lecture. For any future experiment we will run small pilots with two or three people to try and avoid similar problems in the future. We also had a problem with the last question in the questionnaire "How do you feel the voices affected your choice of character?" as some users interpreted that as a general question of how voices affect opinions on people.

4.3 Results and analysis

This section presents the results from all the experiments. The results are presented by comparing the results from each of the three experiments for each of the attributes shown in Table 3.

The results of each test were analysed using the non-parametric Friedman two-way test in order to see if the differences observed in the scores of each character in each test were statistically significant. As required by the Friedman test the scales on the questionnaire were given a weighting. The scale was weighted from one to seven, the total scores were calculated for all the users and then these were normalized from 0 to 1.

Friedman analysis tests the null hypothesis that the *k* repeated measures or matched groups come from the same population or populations with the same median. The alternative hypothesis is then that at least one pair of conditions or samples has different medians. Therefore, when the value obtained from a Friedman test is significant, it indicates that at least one of the conditions differs from at least one of the other conditions; in order words, in our case at least one of the average scores of one of the characters is statistically significantly different from the average score of one of the others. The test does not tell us which one of the results is different nor how many of the groups are different from one another. In those cases where the value obtained from a Friedman test was significant, we used a multiple comparison between the groups in order to identify which of the sample were different from one another. In this work we used non-parametric statistics since it was not possible to assume that the scores under analysis were drawn from a population distributed in a certain way (distribution-free).

In order to find the degree of confidence for the results for the choice of character we used the chi-square goodness-of-fit method, which is normally utilized to analyse the number of subjects, objects or responses, which fall into various categories. See Appendix 5 for more detailed information.

4.3.1 Face Analysis

To analyse the results of all these three experiments, the scales on the questionnaire were given a weighting. The scale was weighted from one to seven, the scores were totalled up for all the users and then these were normalized from 0 to 1. The results of each test were analysed using the non-parametric Friedman two-way test (see section 3.4.3), in order to see if the differences observed in the scores of each character in each test were statistically significant.

4.3.1.1 Friendliness:

For this characteristic, two distinct patterns in the three experiments were clearly observed:

1) In the one and two-way interaction experiments, it can clearly be seen that the friendliness rating was proportional to the degree of abstraction of the character, the friendliest being the most abstract character (character A), then the cartoon character (character B) and finally the least friendly character being the least abstract character (character C) (see Figure 4.4 and Figure 4.5). In these two experiments the differences observed between the average score of each character was statistically significant with a confidence of more that 95%, except for the difference between character B and C in the non-interactive experiment where the confidence level was only 75%.
2) In the two-way experiment characters A and B were rated equal and both of them were significantly friendlier than character C (see Figure 4.6). In this experiment the difference between character C compared to A and B was statistically significant with a confidence level greater than 90%.

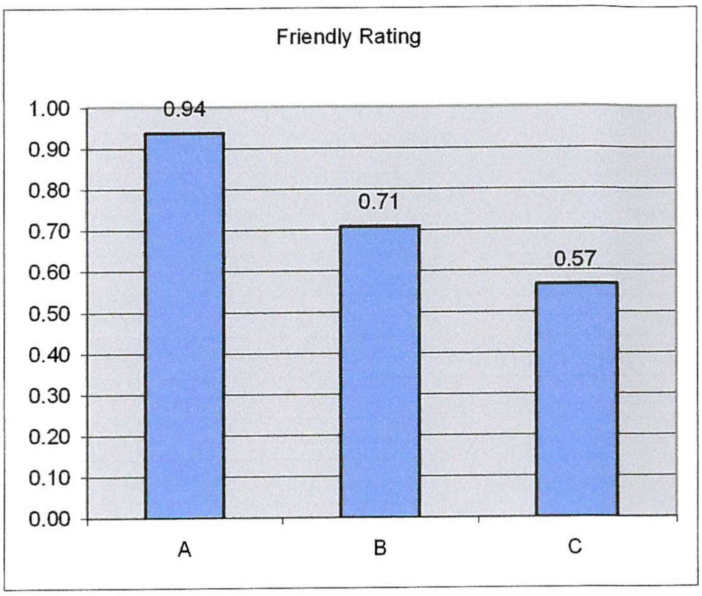


Figure 4.4: Shows the normalized scores for Friendliness in the no-interaction experiment.

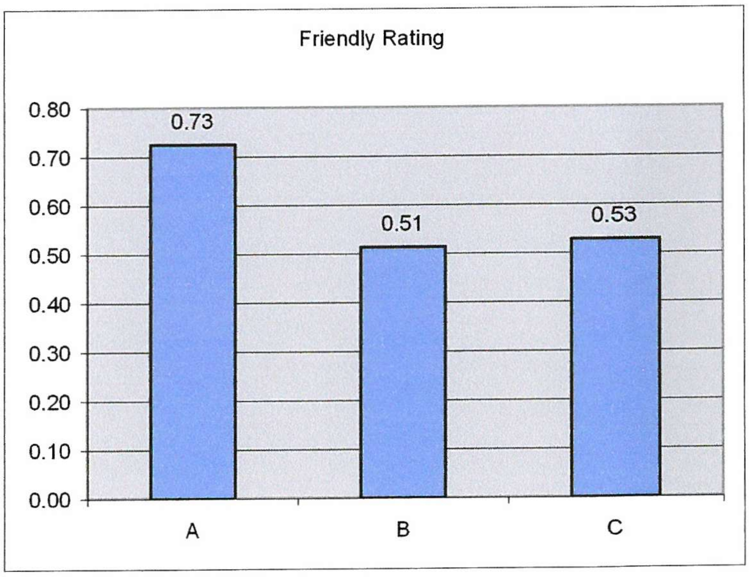


Figure 4.5: Shows the normalized scores for Friendliness in the one-way-interaction condition.

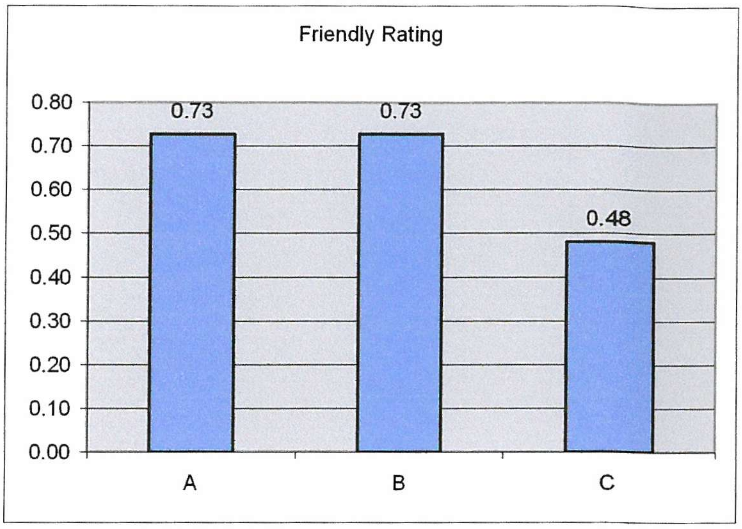


Figure 4.6: Shows the normalized scores for Friendliness the two-way-interaction experiment.

It is important to observe that in the two-way-interaction experiment the cartoon and the smiley characters were rated equally, while in the other experiments (no-interaction and one-way-interaction) the degree of friendliness was always proportional to the degree of abstraction of the characters.

The direct comparison between the scores obtained for the three experimental conditions are tabulated in Table 4.6.

	A	B	C
No Interaction	0.94	0.71	0.57
One-way Interaction	0.73	0.51	0.53
Two-way Interaction	0.73	0.73	0.48

Table 4.6: Shows the normalized scores for Friendliness in all the experimental conditions (no-interaction, one-way interaction and two-way interaction).

Character A scored higher friendliness levels in the no-interaction experiment, 0.94. The users scored this character less, to 0.73, during the two interactive tests (one-way and two-way interaction).

In the case of character B, the no-interaction experiment results in a score of 0.71, which decreased in the one-way-interaction experiment to a value of 0.51, and then significantly increased by the two-way-interaction experiment to its highest value of 0.73. This behaviour of the one and two-way interaction test (i.e. reduction followed by an increase in the score) was frequently observed in the other characteristics evaluated in this work.

Finally the degree of friendliness of character C was always inversely proportional to the order of interaction of the test, with a score of 0.57 with no interaction and 0.48 in the case of two-way interaction, suggesting that the users became more disappointed with the character the more they got to know it.

One factor that remained constant throughout the experiments is that the realistic characters were thought of as being less friendly than the more abstract characters, this is in accordance with McCloud's (McCloud, 1993) and Thorison's (Thorison 1996) proposals. The author considers that a reason might be that the realistic computer generated characters might be intimidating users. This is supported by the following comments made by the users:

- The following comments were given about the female character used for the first and second experiments (two-way and no interaction):
  - *"was scary"*
  - *"scares me"*
  - *"is very scary!"*
  - *"is SCARY!!!!!! She's staring at me. She has a pimple. I think she is laughing at me."*
  - *"C is evil"*
- While regarding male character used for the third experiment (one-way interaction); the following comments were given:
  - *"is far too scary..."*
  - *"looks neurotic"*

The author has considered a few possibilities for why this might be happening, after the first two experiments (two-way and no interaction) it has also been considered that it could either be that particular character was scaring the users, however the alternative character used for the one way interaction experiment was also described as scary. It was also considered that users might be expecting realistic looking characters to move in a less cartoon like fashion and more like humans, if this was the case it was something that would be improved by using the better animated character. This idea that users expect more from the more realistic characters can be seen in a remark from one of the users when commenting about character C (see Figure 2) during the one-way interaction experiment:

- *"doesn't behave like a real human, eye movements, you get confused"*

However, even though the other two more abstract characters did not "behave like a real human", no comments were made about behaving like humans as the users did not expect them to behave as such.

4.3.1.2 Intelligence:

Another pattern identified was that the less abstract a character was the more *intelligence* the users will attribute to it. This result was repeated in all three experimental conditions (see Figure 4.7, Figure 4.8, Figure 4.9 and Table 4.6) and it replicates Koda's results (Koda & Maes, 1996). I think perhaps that apparent intelligence is not necessarily something that designers of anthropomorphic conversational interfaces should aim for. If the characters apparent intelligence greatly surpasses its capabilities then users are likely to be disillusioned when their expectations are not fulfilled.

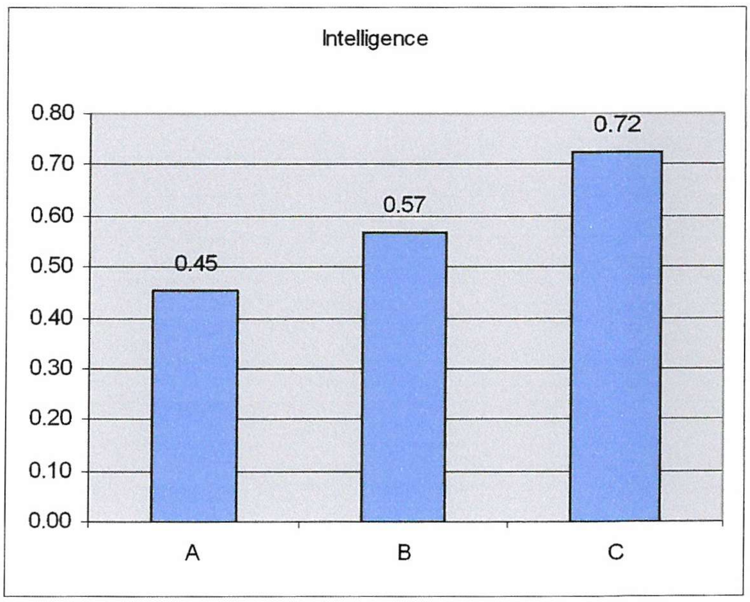


Figure 4.7: The normalized scores for Intelligence in the no-interaction condition.

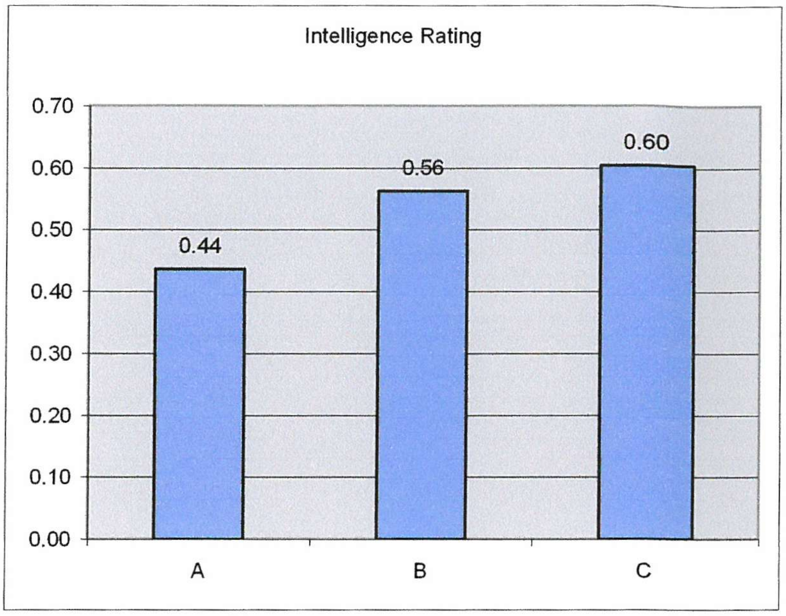


Figure 4.8: The normalized scores for Intelligence in the one-way-interaction condition.

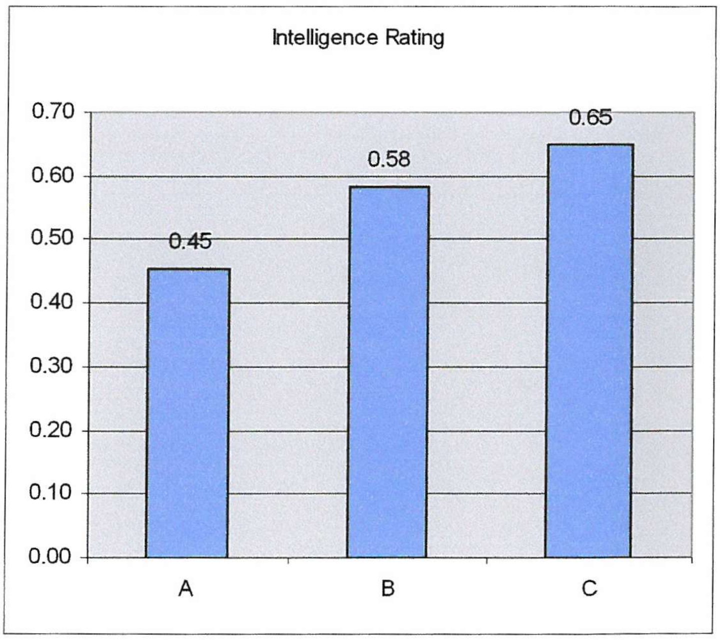


Figure 4.9: The normalized scores for Intelligence in the two-way-interaction condition.

	A	B	C
No Interaction	0.45	0.57	0.72
One-way Interaction	0.44	0.56	0.60
Two-way Interaction	0.45	0.58	0.65

Table 4.7: The normalized scores for Intelligence in all the experimental conditions (no-interaction, one-way interaction and two-way interaction)

It was anticipated that users would see the more realistic characters as more intelligent since they seem more human and less like objects. Intelligence being a specifically human characteristic it is understandable that users see the more human like characters as more intelligent.

Although it appears that the differences between the average *intelligence* scores of each of the characters was well defined, it was observed from the non-parametric tests carried out during the evaluation of the results, that only in the no-interaction experiment is the difference in the scores of any statistical significance with a confidence always greater than 85%, which is not enough to reach a firm conclusion but it indicates that there might be an effect worth investigating further.

In the one-way interaction experiment the difference between character B and C was not statistically significant and the differences between A and B and A and C were significant with a confidence level greater than 95%. On the other hand, in the two-way interaction experiment only the difference between characters A and C was found to be statistically significant with a confidence greater than 90%.

In these tests no significant effects from the degree of interaction were observed on the score obtained by each character. Each character had almost the same score in each experiment except character C in the no-interaction experiment, where it scores it's highest value (see Table 4.7).

4.3.1.3 Pleasantness

A pattern was also identified with the pleasantness attribute that was a similar pattern to the one observed for friendliness. This is where pleasantness is inversely proportional to the level of abstraction. We observed this pattern in all three experimental conditions (see Figure 4.10, Figure 4.11, Figure 4.12 and Table 4.8).

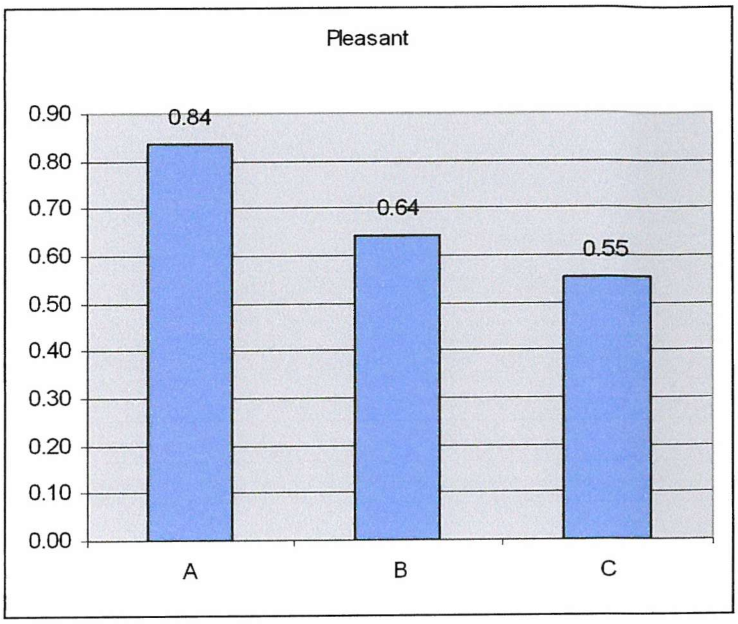


Figure 4.10: The normalized scores for Pleasantness in the no-interaction condition.

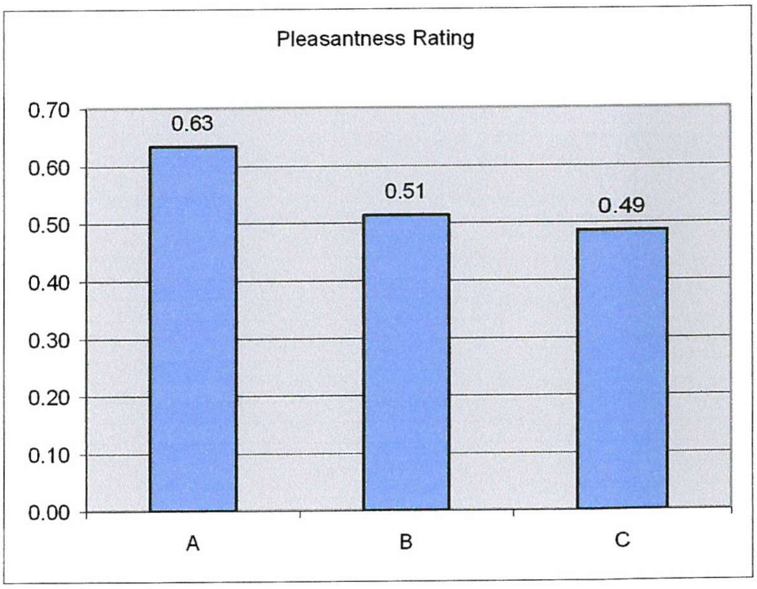


Figure 4.11: The normalized scores for Pleasantness in the one-way-interaction condition.

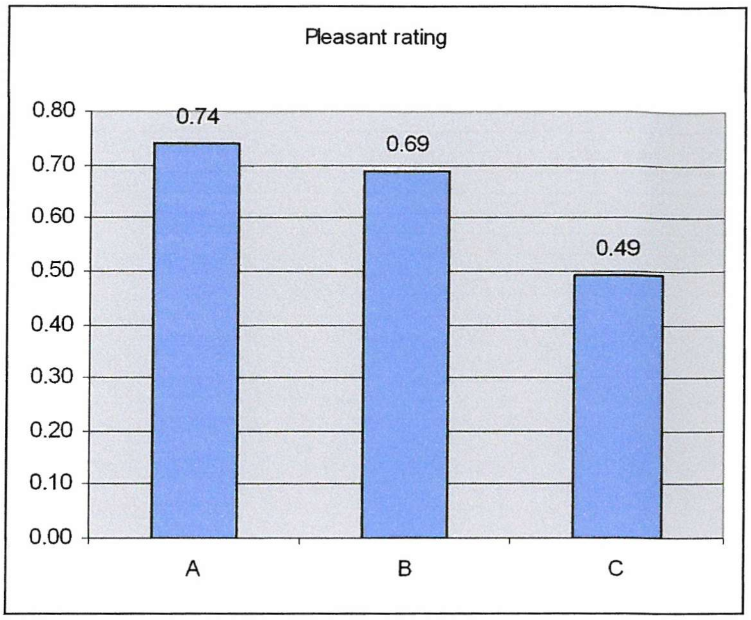


Figure 4.12: The normalized scores for Pleasantness in the two-way-interaction condition.

	A	B	C
No Interaction	0.84	0.64	0.55
One-way Interaction	0.63	0.51	0.49
Two-way Interaction	0.74	0.69	0.49

Table 4.8: The normalized scores for Pleasantness in all the experimental conditions (no-interaction, one-way interaction and two-way interaction).

Pleasantness, in the Oxford English Dictionary, is defined as the quality of being pleasant. Pleasant is in turn defined as: having pleasing manners, demeanour, or aspect; agreeable, cheerful, good-humoured. On reflection pleasantness and friendliness are perhaps quite closely related. This association between *friendliness* and *pleasantness* can be clearly seen in Table 4.6 and Table 4.8 as they follow similar patterns. Another similarity to the friendliness attribute is that, in the one-way interaction experiment the users gave this attributes its lowest scores. However, the score given by the users increases during the two-way interaction experiment (see Table 4.8), showing again that it might be better to not have any interaction with the characters if the only interaction possible is merely partial. Once again the overall pattern of the scores is preserved across the different conditions.

The author also considered that it is likely that the more abstract characters leave more open to the imagination and thus are less likely to be disliked. McCloud (McCloud, 1993) argues that when viewers see cartoon characters they see a reflection of themselves. That cartoons are like an empty shell that enables us to not just watch the cartoon, but to "become it".

4.3.1.4 Interestingness

We have identified that the highly abstract character was classed as being less interesting to look at than the less abstract ones. In all three experimental conditions the abstract character is rated as considerably less interesting than the others (see Figure 4.13, Figure 4.14, Figure 4.15 and Table 4.9).

In this case the *Interesting* scores obtained for each character in each of the experiments were very similar. However, as with other attributes, the users gave the *interesting* attribute its lowest score in the one-way interaction experiment.

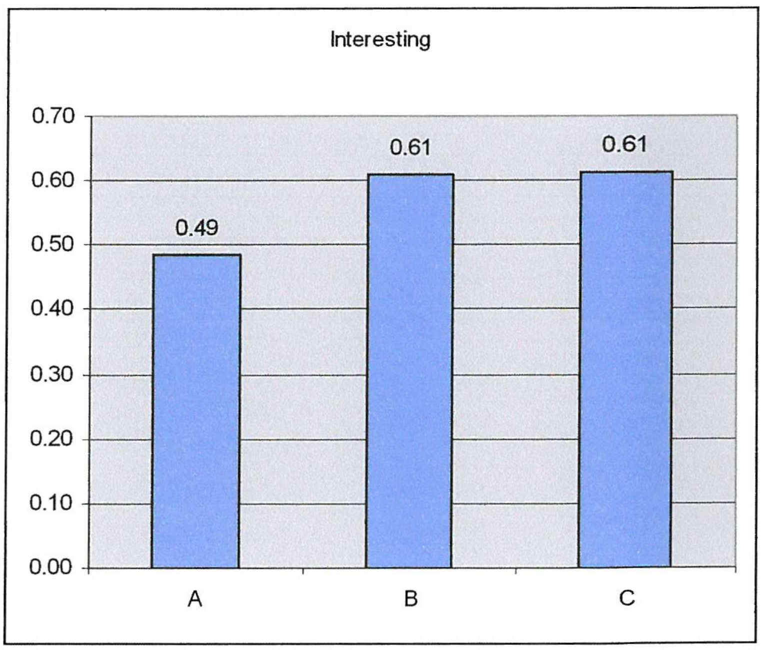


Figure 4.13: The normalized scores for the interesting rating in the no-interaction condition.

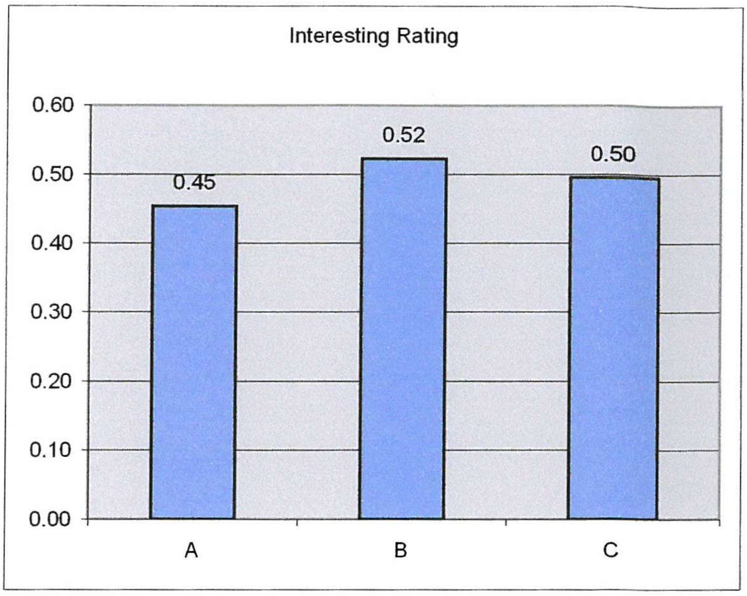


Figure 4.14: The normalized scores for the interesting rating in the one-way-interaction condition.

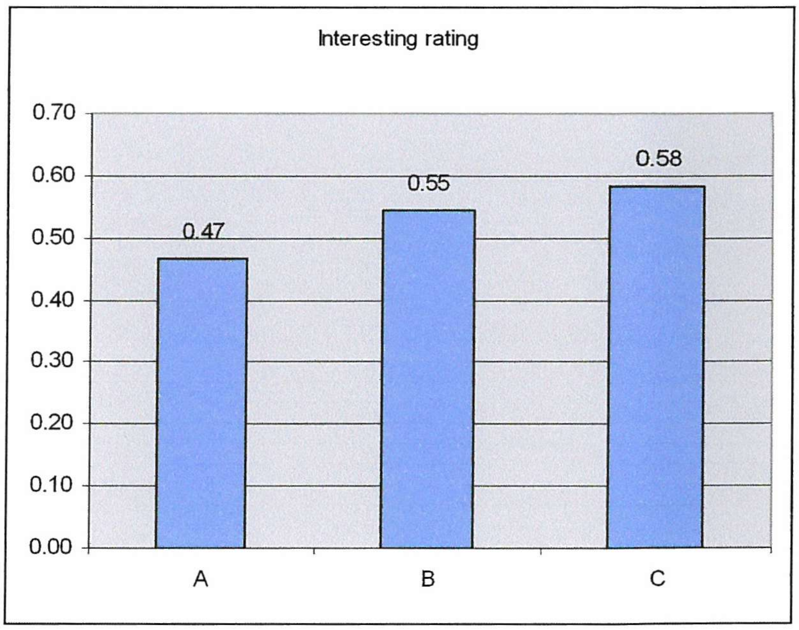


Figure 4.15: The normalized scores for the interesting rating in the two-way-interaction condition.

	A	B	C
No Interaction	0.49	0.61	0.61
One-way Interaction	0.45	0.52	0.50
Two-way Interaction	0.47	0.55	0.58

Table 4.9: The normalized scores for Interesting in all the experimental conditions (no-interaction, one-way interaction and two-way interaction).

The non-parametric tests showed that in the no and one-way interaction experiments the differences observed between the characters A and B and A and C were statistically significant with a confidence level of more than 85% only, which makes it worthy of a mention but not enough to reach a firm conclusion. However, the difference between the characters B and C was not statistically significant. However none of the results obtained in the two-way experiment were found to be statistically significant. For this characteristic it appears that users respond in a similar way independently of the experiment carried out.

4.3.1.5 Helpfulness

As explained earlier this characteristic was not evaluated as part of the no-interaction condition experiment due to the inconsistency of this attribute with this type of experiment. It appears from the scores obtained in the two remaining experimental conditions that users attributed slightly more helpfulness to the character B, the cartoon character (see Figure 4.16, Figure 4.17 and Table 10). However, the non-parametric test for all the experiments where *Helpfulness* was evaluated; shows that the difference between the average scores was not statistically significant. It appears that this characteristic is too subjective to be analysed by our experiments.

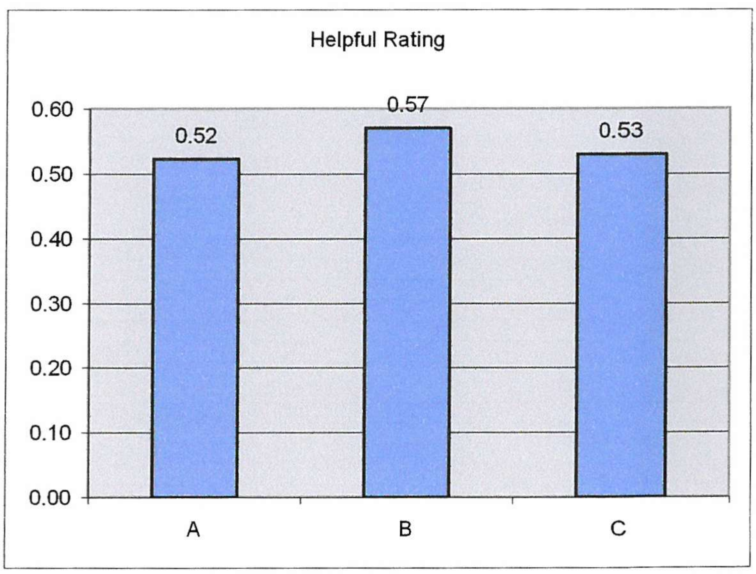


Figure 4.16: The normalized scores for the helpfulness rating in the one-way-interaction condition.

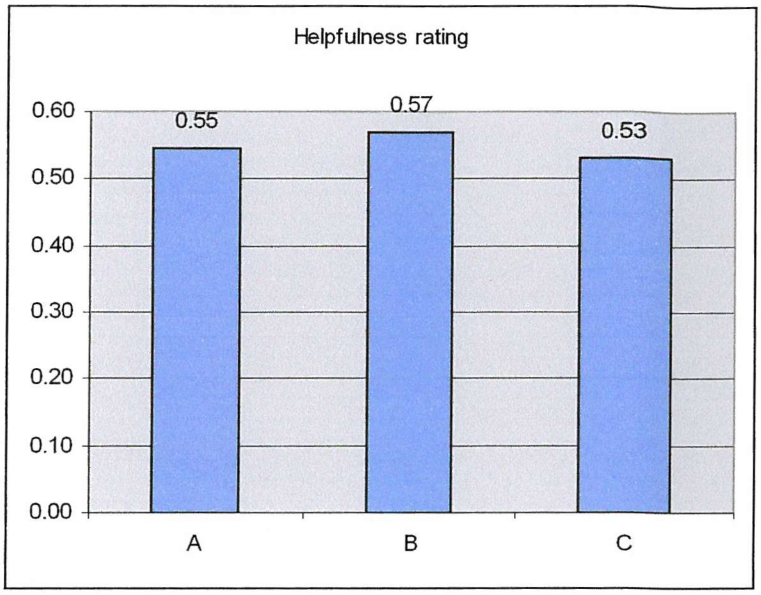


Figure 4.17: The normalized scores for the helpfulness rating in the two-way-interaction condition.

	A	B	C
One-way Interaction	0.52	0.57	0.53
Two-way Interaction	0.55	0.57	0.53

Table 4.10: The normalized scores for Helpfulness in two experimental conditions (one-way interaction and two-way interaction).

4.3.1.6 Choice of character

During the one-way interaction experiment users were asked to choose a favourite character and the overwhelming majority chose face B, the cartoon character (see Figure 4.18). In order to find the degree of confidence for these results the chi-square goodness-of-fit method was utilized, which is normally utilized to analyze the number of subjects, objects or responses, which fall into various categories. This analysis shows that the results given in Figure 4.18 are more than 95% statistically significant.

The overall scores for each of the characters in the one-way interaction experiment (see Figure 4.19) were worked out by taking into account all the characteristics considered in each experiment. These overall scores were surprisingly close considering it was clear that most users preferred the cartoon character. We propose the final success of the cartoon character is because it is a middle of the road character that does not score too low or too high on any particular attribute. On the other hand both the very abstract character and the realistic characters score very high for some features and very low for others. We put forward the theory that the realistic characters are proving to be too "scary" as people are expecting them to act like human beings and not cartoons with text to speech voices. The very abstract character although perceived as being friendly and pleasant, is not seen as being very interesting to look at.

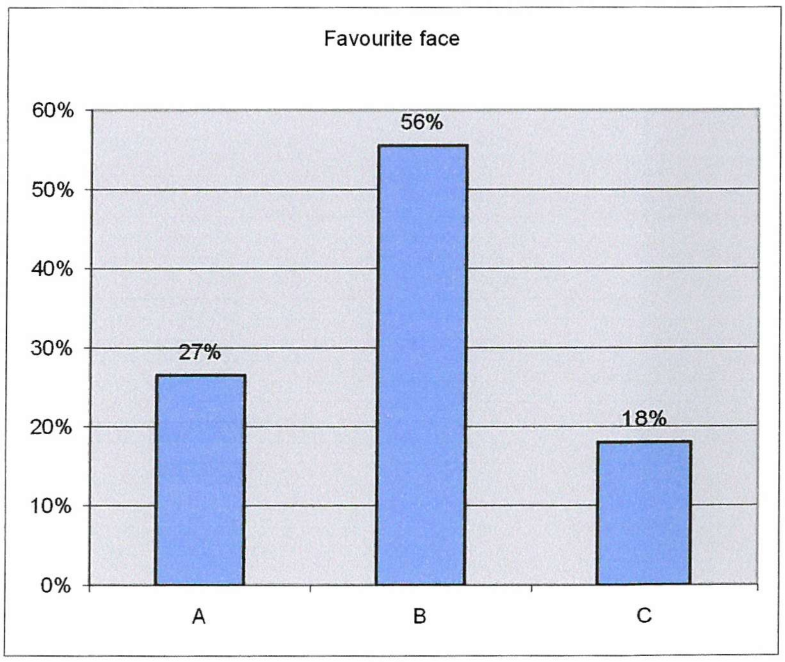


Figure 4.18: Favorite face on the one -way interaction condition as a percentage.

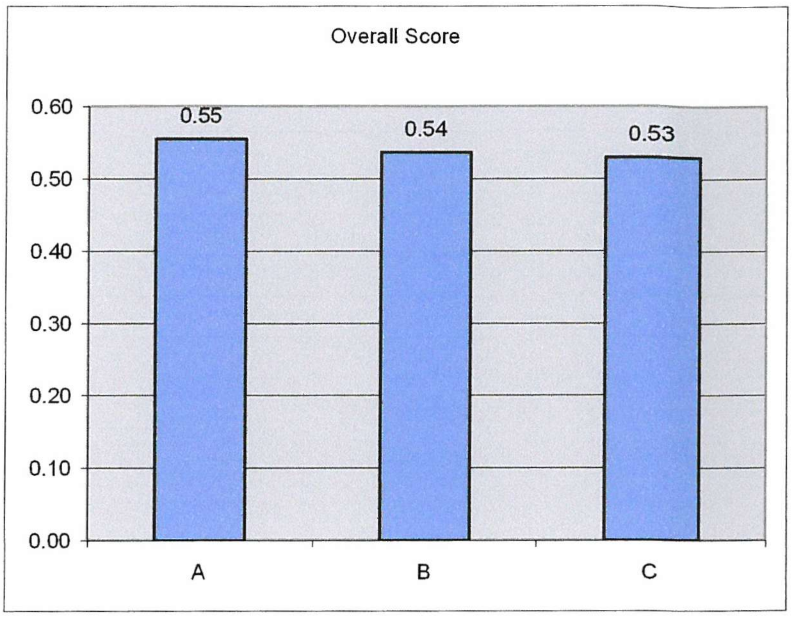


Figure 4.19: The overall normalized score for the three characters in the one-way-interaction condition.

The author considers that the 'scariness' factor of the realistic characters is something that could be overcome if the user expectations are fulfilled by the character's capabilities. These capabilities include artificial intelligence techniques, realistic animation and human like voices, as this is what the user expects when confronted by a very realistic looking character. The author sees this as merely a technological barrier that once broken would allow us to create very realistic believable agents (Mateas, 1997) (Bates, 1994) which people might not find so awkward to interact with.

This is supported by user comments such as:

- *"The more realistic the face the more you are expecting a more realistic voice. When this doesn't happen you are less likely to listen as attentively to what is being said"*

None of the characters (see Figure 1 & 2) were perfect as demonstrated by the user comments:

- *"Face A doesn't really look like a face when talking, just lots of random blobs flashing up. Face C is far too scary..."*
- *"Didn't like the eyes of A & B..."*
- *"Didn't like the mouth movements of face A"*
- *"Eyes on face B look freaky..."*
- *"...face B would have been more appealing without black eyes"*

between the scores for voices 1 and 2 and 2 and 3 were statistically significant, while between 1 and 3 the difference was not significant.

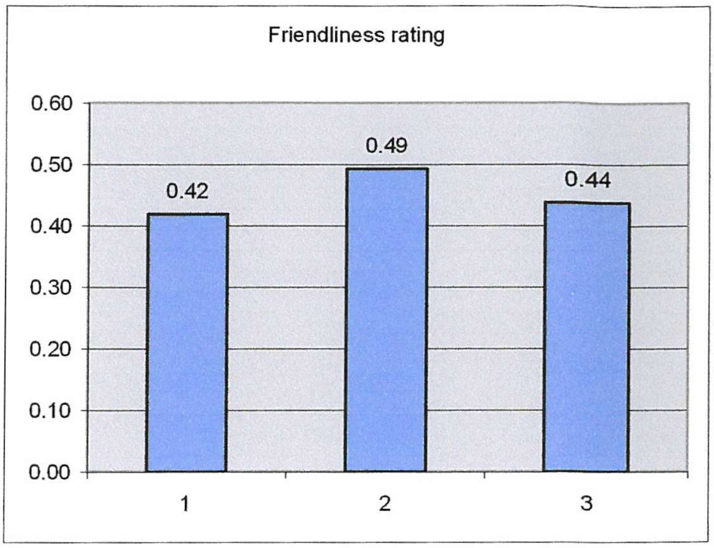
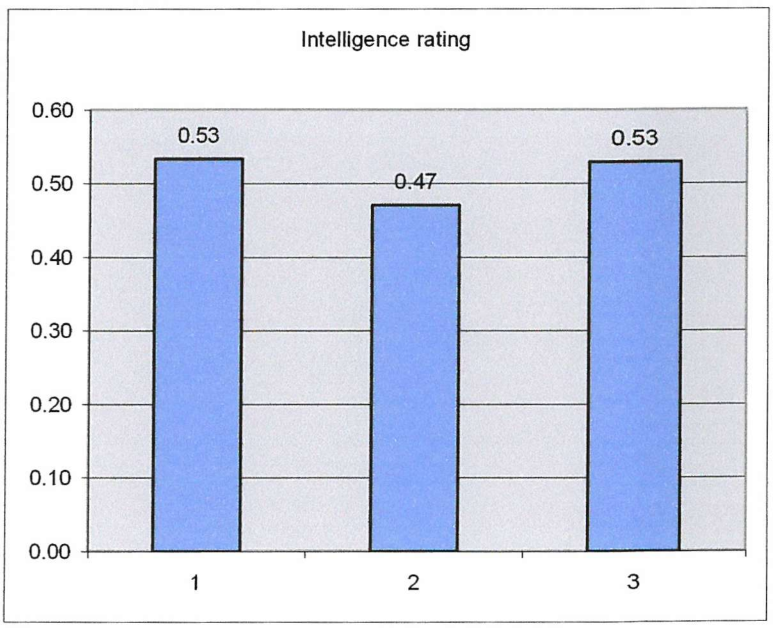


Figure 4.20: Shows the normalized scores for friendliness score for the three voices in the one-way-interaction condition.

4.3.2.2 Intelligence

In contrast with the previous test here the score for voice 2 was rated the lowest, 1 and 3 being rated identically (see Graph 1). As before in this case the only score that was statistically significant was the one corresponding to voice 2.



Graph 1: The normalized scores for intelligence score for the three voices in the one-way-interaction condition.

4.3.2.3 Pleasantness

In this case it appears from the average values that voices 2 and 3 were almost identical and rated with a higher score than voice 1 (see Figure 4.21). From the non-parametric test it was found that the scores obtained by voices 1 and 2 were statistically significant with a confidence level of only 85% and those from 1 and 3 only with a confidence level of 70%, neither of which are significant enough to reach a firm conclusion. The difference between the scores for voices 2 and 3 was not found to be statistically significant.

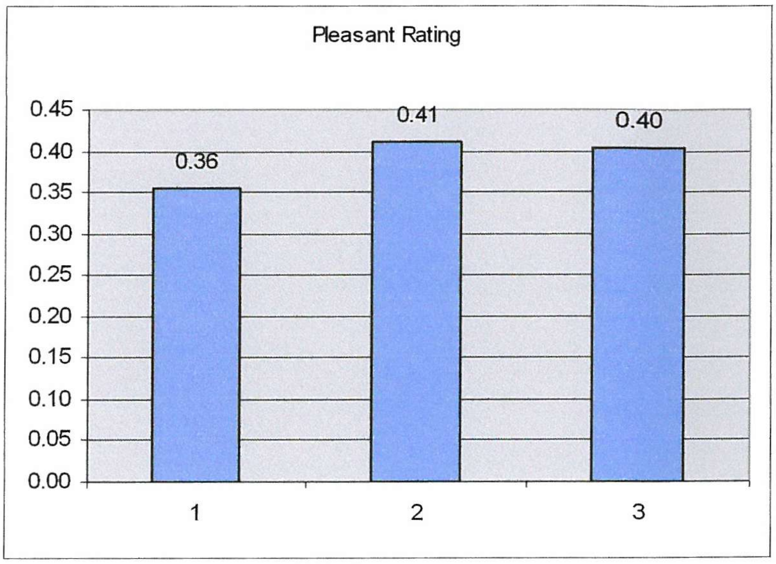


Figure 4.21: The normalized scores for pleasantness score for the three voices in the one-way-interaction condition.

4.3.2.4 Interestingness

This was the only case where the scores between the three voices were found to be statistically significant and with a confidence greater than 95%, except between voices 1 and 3 where the confidence level was only 70%, therefore we cannot conclude that there's any perceived difference between them. The highest score was obtained by voice 2, followed by voice 3 and voice 1 being the lowest (see Figure 4.22).

Figure 4.22: Shows the normalized scores for intelligence score for the three voices in the one-way-interaction condition.

4.3.2.5 Understanding

Although voice 2 was theoretically the most advanced, it scored low in understanding (see Figure 4.23). This could have been due to the voice braking up towards the end of the presentation, as this particular TTS software is very CPU intensive. As in most of the previous cases here only the score obtained for voice 2 was statistically significantly different from the others.

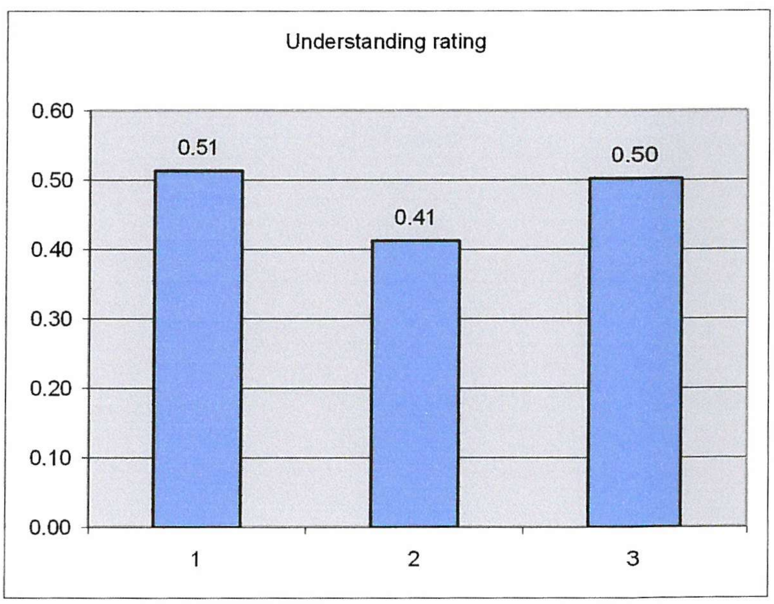


Figure 4.23: The normalized scores for understanding for the three voices in the one-way-interaction condition.

4.3.3 *User strategy*

It was considered that it would be a worthy exercise to explore if there is any link between how much like a human the test subjects treated the characters and the following factors: their favourite character, the level of anthropomorphism displayed by the character, the order in which the interaction took place and all the criteria examined in the two way interaction experiment.

To quantify how much like a human users treated the characters, the author devised a method where input from the test subjects is categorized in two categories. One reflecting the sort of queries that you would expect a human to make to another human and the other the sort of query you would expect a human to make to a robot. The rules we used to differentiate from these two types of queries are shown in Table 4.11.

Type of user input which indicates that the Agent being perceived as a Search engine or Robot	**Types of user input which indicates that the Agent is being perceived as a human or at least intelligent enough to understand more complex commands**
• Keywords • Simple direct commands or queries. • Insults	• Questions • Complex commands or queries • Any natural language input that is not a set of keywords or a simple command or query. For example chitchat.

Table 4.11: Categories employed to work out the 'rudeness rating' of a set of user queries.

Keywords: We define keywords as any one word or group of words that does not form a complete sentence.

E.g.:

- *Cars*
- *Seaside villa*

Questions: It was thought that a user who posed a question to the character would be assuming a certain amount of knowledge from the character and thus this user would be treating the character like a human.

E.g.:

- *What is the offside rule?*
- *Where did Elvis live?*

Simple commands/queries and complex commands/queries: We define a simple direct command or query as one that has a simple target without conditions or unnecessary additional words used to make the conversation more polite.

E.g.:

- *"Find a website about cars."*

  This is a direct command.

- *"Could you please find me a website about cars."*

  This is not a simple direct command as the user was polite to the character by saying "Could you please".

- *"Find me a website with technical information about the Corvette C5R Le-Mans 24 hour winner."*

This is not a simple direct command as the user has specified a condition to the query "technical information" and has provided extra information "Le Mans 24 hour winner" which the user has assumed the character can understand and thus is assuming a certain amount of intelligence.

Insults: An insult is anything considered to be rude too be said to a stranger under similar conditions to the experiment, or in other words, anything the author would find shocking or offensive if any of the test subjects were to say it to him. Originally, the category insult was not included as it was assumed that test subjects would not resort to insulting the character during these short experiments, however one particular user made queries, which were considered to be rude, and thus the insult category was created. This user made the queries *"who [inveted] hypertext then you squared jawed surfer boy."* and *"My god you're ugly."*. The author would be offended or at the very least rather shocked if a test subject he had never seen before called him *"a square jawed surfer boy"* or *"ugly"* so these queries were qualified as an insult.

Using these categories the ratio of 'rude' queries to the total number of queries is calculated. We called this the 'rudeness rating'. The rudeness ratings for all three characters were calculated. For a complete listing of queries and how they were categorized see Appendix 6.

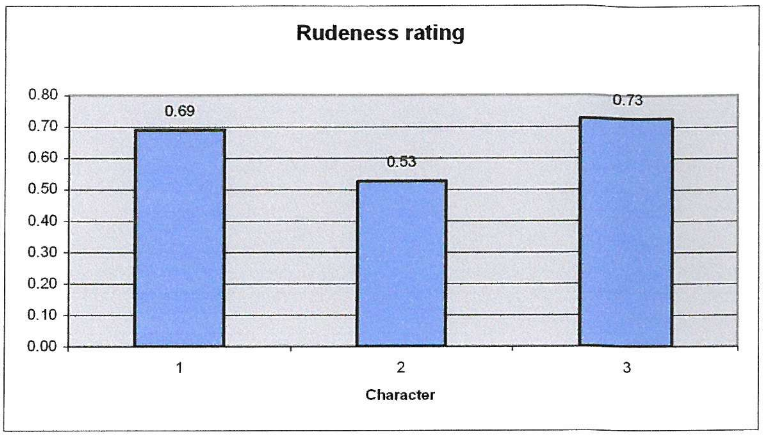


Figure 4.24: Rudeness rating for characters A, B and C respectively.

It is important to note that the data shown in Figure 4.24 appears to be the reverse of the data displayed on Figure 4.18, which shows users who picked a particular face as their favorite on the one -way interaction experiment as a percentage. So perhaps if instead of a rudeness rating we had a politeness rating then we would obtain a graph, which matches the pattern shown on Figure 4.18 (see Figure 4.25).

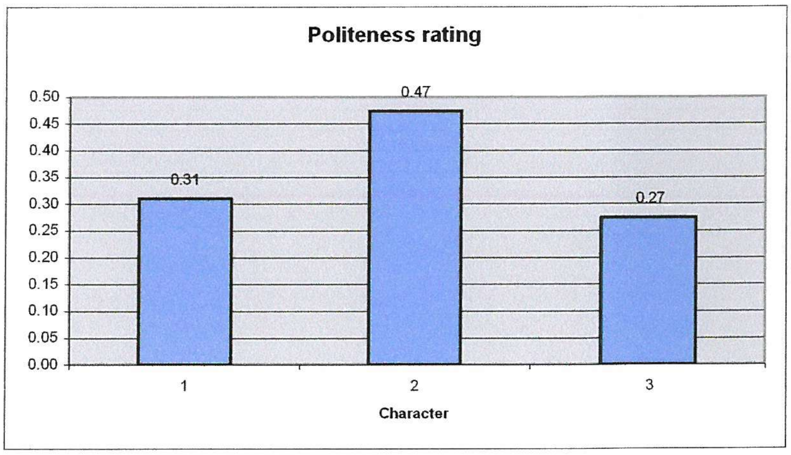


Figure 4.25: Politeness rating for characters A, B and C respectively.

Therefore it would appear that users treat their favourite characters more politely than those characters, which they do not like as much. It is also worth noting that all the 'rudeness ratings' were high for all characters as no rudeness at all would be a zero and these values were all above 0.5 and one was 0.73, which is close to complete rudeness (1.0).

The author considers the approach described here as worthy of further development. A fully automated system developed using this method could prove a useful method for testing large-scale conversational interface acceptance without implicitly designing user experiments. This would reduce costs incurred by companies and institutions releasing these interfaces.

4.3.3.1 Summary

We have made the following observations during the course of our experiment:

The realistic faces were seen as less friendly and even scary: We propose the *scary* factor is due to the character's behaviour not matching the user's expectations of the human looking face. We put forward that this is a technological barrier that will be overcome with time.

The more abstract a character the more friendly/pleasant it seems: It is possible that less realistic more abstract characters leave more open to the imagination and thus are less likely to be disliked. The *friendliness/pleasantness* attributes are inversely proportional to the level of abstraction.

The more abstract the character the less intelligence users will attribute to it: It was anticipated that users would see the more realistic characters as more intelligent since they seem more human and less like objects. Intelligence being a specifically human characteristic it is understandable that users see the more human like characters as more intelligent.

The more abstract the character the less interesting it is to look at: It is not surprising that users rate the simpler characters as less interesting to look at than more complex ones.

The users favoured a moderately abstract character: The cartoon character is the users favourite even though this is not reflected by higher overall scores than the other two. We propose that its success is because it scores well in all characteristics as opposed to the other characters, which score very high for some features and very low for others.

Although partial interaction does lower the scores attributed by the users it does not appear to change the overall patterns displayed. Therefore when developing future studies the use of partial interaction could prove a useful tool in developing quicker experiments.

On the issue of voice, subjects were not always aware that different voices were employed during the experiment. This may be because subjects were focusing on other aspects of the interaction instead of the particular voice used. It is also worth mentioning that voice 2 scored lowest in understanding. It is thought this may be due to the voice skipping, as the laptop being used for the experiment sometimes could not cope with that particular TTS engine. It appears that tailoring the software to the hardware being used is the key to the success of the interface and the particular sound of the TTS being used might be relatively unimportant.

The experimental results indicate that users treat their favourite characters more politely than characters they do not like as much. This finding could be exploited to build a system which analyses user and agent interaction to determine anthropomorphic agent acceptance without carrying out user experiments.

The review of the papers detailing the no-way and two-way interaction experiments (Power, Will et al. 2002) (Power, Damper et al. 2002) highlighted a problem with the characters chosen, it was pointed out that having a mixture of male and female characters added an extra variable which we weren't accounting for.

5 Gestures

This chapter explores how facial gestures affect the users' perception of humanity exhibited by the agent. User trials were performed to test whether adding gesticulative ability to a character affects the way users perceive the character. To accomplish this a storytelling setting with agents of two different levels of gesticulative ability was utilized:

1. No gesticulative ability: The agent exhibits a neutral expression throughout the experiment.
2. Some gesticulative ability: The agent still exhibits a neutral expression but is capable of blinking, nodding, head shaking and glancing at areas of the screen.

Throughout the experiment the agent displays a neutral emotionless expression, it does however show emotion twice for each of the aforementioned two cases. The agent smiles once when introducing himself and once again when saying good-bye to the user. This limited show of emotion was deemed necessary as there was a risk of making the agent seem to harsh and thus alienate or put off users.

We are unable to conclude what sorts of effect gestures have on users, however, we do conclude that when an anthropomorphic agent is presenting the user with some sort of graphical information users tend to look at this information and not at the agent itself. Prompting us to ask ourselves the questions; when are anthropomorphic agents necessary? And even, are they necessary at all?

5.1 Introduction

The aim of this chapter is to investigate users' reaction and assumptions when interacting with anthropomorphic agents and how these assumptions and reactions are affected by gestures. A storyteller setting was used as a test bed for this experiment. Two groups of test subjects were compared, one group exposed to a storyteller capable of displaying gestures and a control group exposed to a character incapable of displaying gestures.

Cassell (Cassell and Thórisson 1999) has found that users rate agents displaying nod and glance gestures more highly than agents with no gestures or agents with emotional facial expressions. Cassell mentions possible flaws in the experimental procedure and subjects that further research of the effect of "nod and gland" is required, however she does conclude that "envelope feedback" (gestures not directly related to the content of the speech) may allow us to create anthropomorphic agents that work better and are better accepted by users. As such we devised an experiment to try and corroborate and expand upon Cassell's findings.

Similarly to our previous set of experiments a set of four qualification tests (friendliness, intelligence, pleasantness and interestingness) was carried out on two user groups to evaluate the effect that gestures had on user's opinions and expectations of anthropomorphic agents.

The experiment was carried out with characters animated using Microsoft Agent (Microsoft 1999) within a PowerPoint presentation in a similar fashion to the one-way-interaction experiment in Section 4.2.5.

Taking into account complaints we had during the two-way-interaction experiment in the previous chapter when the character giving the lesson was described as being boring and monotone, it was decided to design a more entertaining setting for this experiment. The storyteller setting was chosen as it was thought that it was a setting that would not bore users, to further enhance user interest slides of the fable, Aesop's the Fox and the Grapes, were included to the PowerPoint presentation.

1.1 Experimental Design

This section describes the experiment carried out in detail.

5.1.1 Experimental setup

Given previous results (see Chapter 4) where there was little difference in results between experiments of different levels of interactions we concluded that a storytelling setting with limited interaction was suitable for this evaluation. Aesop's "The Fox and the Grapes" (see Figure 1) has been chosen as the story to be related to the listener.

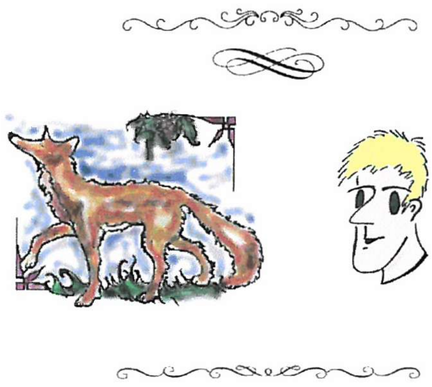


Figure 5.1: A slide from the presentation showing the storyteller and the fable illustration.

5.1.1.1 Users

The experiment was advertised locally across the University. We obtained 25 test subjects divided into 2 groups, of 12 and 13 subjects respectively (see Table 5.1 and Table 5.2). Individuals from group A had the story presented to them by the agent with limited gesticulative ability whereas individuals from group B had the story presented to them by the agent with more advanced gesticulative ability.

Number of Users	13
Gender distribution	5 male and 8 female
Age distribution	Age group 20-37

Table 5.1: User profile for the first group.

Number of Users	12
Gender distribution	9 male and 3 female
Age distribution	Age group 19-30

Table 5.2: User profile for the second group.

5.1.1.2 Presentation

The presentation consisted of the user being presented with a title slide with an arrow labeled "Next" for them to click on (see Figure 5.2).

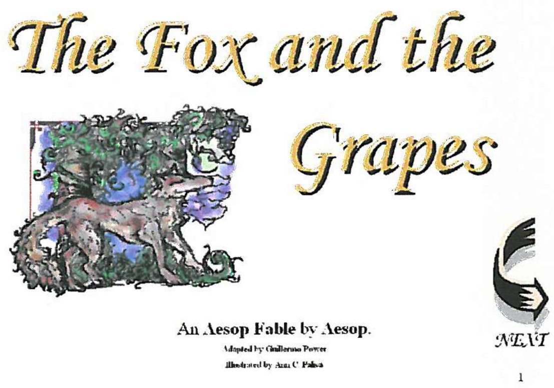


Figure 5.2: The first slide of the presentation.

Following that slide they are presented with another slide containing instructions on how to 'wake up' the storyteller (see Figure 5.3).

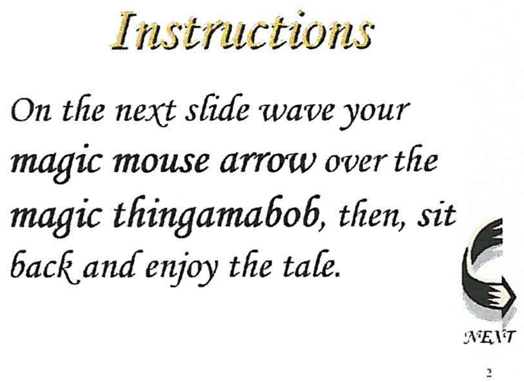


Figure 5.3: Shows the slides with instructions on how to wake up the storyteller.

On the next slide the users are presented with a "Thingamabob" (see Figure 5.4). The "Thingamabob" is a picture of "staff of Ra Headpiece" from the film "Indiana Jones and Raiders of the Lost Ark".

Figure 5.4: Shows the magic Thingamabob.

When the user waves the mouse arrow in front of it the storyteller literally makes an explosive appearance (see Figure 5.5).

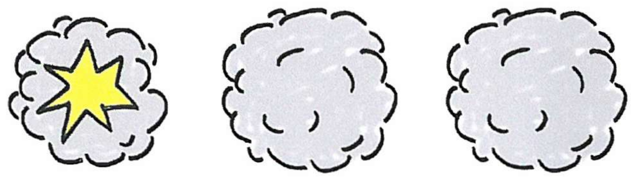


Figure 5.5: Shows the frames of animation used for the explosion.

After the storyteller makes an appearance he thanks the user for waking him up, jokingly says that he borrowed the "Thingamabob" from Indiana Jones and goes on to introduce the fable he is about to narrate to the user and then proceeds to relate this fable across 5 slides (see Figure 5.6).

Figure 5.6: Shows the story slides, which are presented by the storyteller.

The transcript of the text related by the storyteller is given in Table 5.3.

Slides 1 and 2	No commentary from the storyteller.
Slide 3	*Hi.* *Thanks for waking me up.* *As a reward I'll tell you a story.* *Interesting thingamabob. Don't you think? I borrowed it from Indiana*
Slide 4	*One afternoon a fox was walking through the forest and spotted a bunch of grapes hanging from over a lofty branch. "Just the thing to quench my thirst" the fox said.*
Slide 5	*Taking a few steps back, the fox jumped and just missed the hanging grapes.*
Slide 6	*Again the fox took a few paces back and tried to reach them but failed. Finally giving up.*
Slide 7	*The fox turned up his nose and said, "They're probably sour anyway," and proceeded to walk away. The moral of this story is.*
Slide 8	*It's easy to despise what you cannot have. See you later, bye bye.*

Table 5.3: Transcript of story related by the storyteller.

The text is exactly the same for both presentations, with facial expressions and without, shown to the test subjects.

5.1.1.3 Gesticulative ability of the storyteller

The storyteller exploits a number of gestures with the aim to enhance communication of the story with the listener. The character used performed the following gestures during this experiment:

- Speaking whilst exhibiting a neutral expression.
- Speaking whilst smiling.
- Deictic gestures like:
  - Quickly glancing at an illustration.
  - Gesturing towards the illustration whilst speaking and smiling.
  - Gesturing towards the illustration whilst speaking and exhibiting neutral expression.
- Nodding.
- Shaking head.
- Quick smile.

- Shaking head.
- Quick smile.
- Blinking.

The frames of animation used can be seen on Figures below:

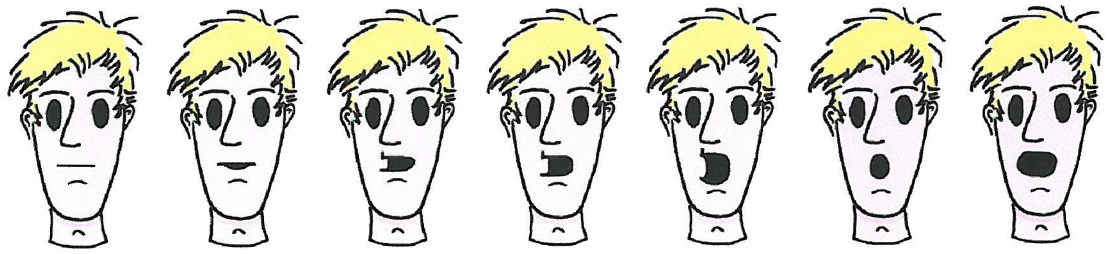


Figure 5.7: Frames of animation used for normal speech, i.e. neutral expression.

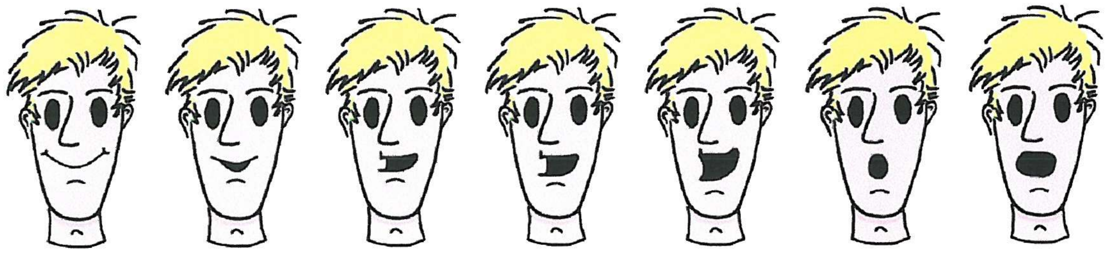


Figure 5.8: Frames of animation used for speech whilst smiling.

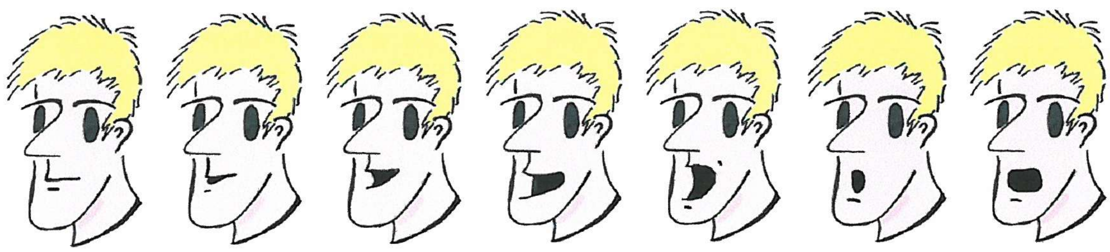


Figure 5.9: Frames of animation used the deictic gesture of looking at the picture depicting the fable whilst speaking with a neutral expression.

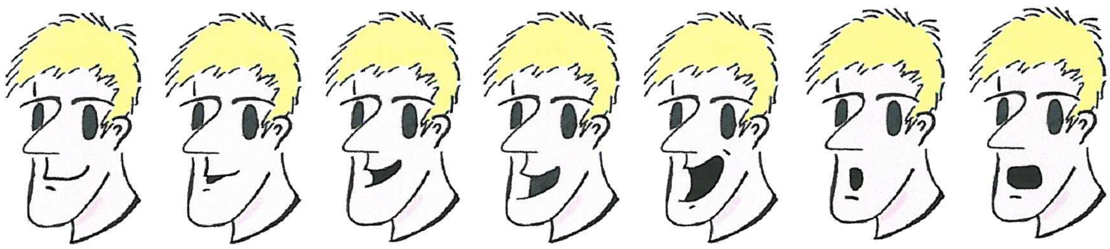


Figure 5.10: Frames of animation used the deictic gesture of looking at the picture depicting the fable whilst speaking whilst smiling.

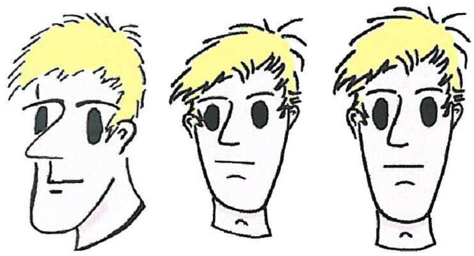


Figure 5.11: Frames of animation used for deictic gesture of glancing at the picture depicting the fable.

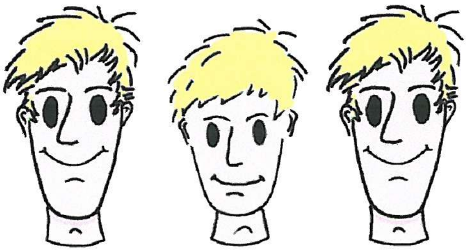


Figure 5.12: Frames of animation used for nodding.

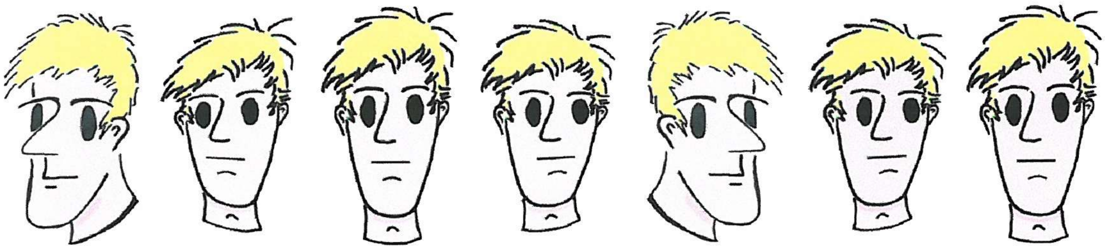


Figure 5.13: Frames of animation used for shaking of the head.

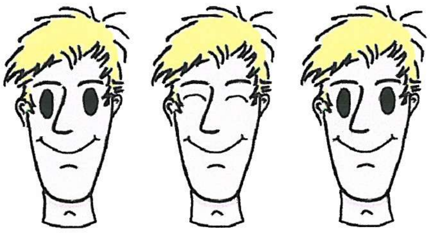


Figure 5.14: Frames of animation used for a quick smile.

The list of gestures exhibited in the presentation is shown in Table 5.4 below.

Gesture / Animation	When this occurs
Quick Smile	During slide 3 after the storyteller jokingly says: *"I borrowed it from Indiana Jones"*.
	During slide 8 After the storyteller nods after saying *"See you later, bye bye"*. Then he proceeds to disappear.
Gesture towards the illustration whilst speaking whilst exhibiting neutral expression.	During slide 3 when the storyteller says: *"Interesting thingamabob"*.
	During slide 5 when the storyteller says: *"Taking a few steps back, the fox jumped and just missed the hanging grapes"*.
Quickly glancing at illustration	During slide 6 after the storyteller says: *"Finally giving up"*.
Gesture towards the illustration whilst speaking and smiling.	During slide 8 when the storyteller says: *"It's easy to despise what you cannot have"*.
Nod	During slide 8 after the storyteller says *"bye bye"*.
Shaking head	During slide 3 after the explosion the storyteller shakes his head as though trying to recover after being stunned by it.
Blink	During slide 1 it happens after the storyteller says: *"Thanks for waking me up"*.
	Then it occurs a further 5 times before every change of slide.
	It is worth noting that every quick smile animation also contains a blinking frame.

Table 5.4: Gestures exhibited by the character.

5.2 Results and analysis

This section presents the results from the gestures experiment. The results are presented in order of attributes examined.

To analyse the results, we once again gave the scales on the questionnaire a weighting. The scale was weighted from one to seven, the scores were totalled up for all the users and then these were normalized from 0 to 1. The results of each test were analysed using the non-parametric Mann-Whitney (see Section 3.4.2) test in order to see if the differences observed in the scores between the two groups of test subjects were statistically significant.

5.2.1 Attributes

None of the results obtained was statistically significant according to the Mann-Whitney analysis carried out. Statistical analysis and user feedback revealed that the majority of test subjects could not tell the difference between the storytellers as most of them were concentrating on the slides presented instead of the storyteller's face (see section 5.2.2) thus results were very similar for both storytellers. The attributes examined are as follows:

5.2.1.1 Friendliness

Users were asked to rank the friendliness of the character on a scale of one to seven from unfriendly to friendly. The normalized scores for friendliness were 0.63 for group 1 and 0.64 for group 2, see Figure 5.16. These are extremely close like some of the other scores for the other attributes in this experiment.

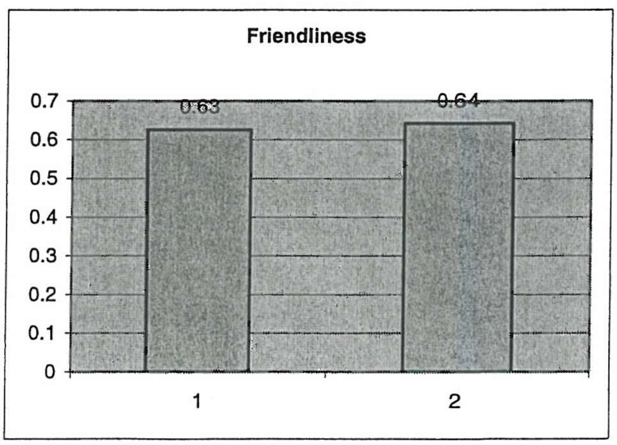


Figure 5.16: The normalized scores for friendliness in the gestures experiment.

Table 5.5 shows N (the number of users), the mean of the rank as explained in and the sum of the ranks (see Section 3.4.2). Table 5.6 shows the results of the Mann-Whitney analysis showing that the results are not statistically significant; see Section 3.4.2 for further information.

Ranks

	group	N	Mean Rank	Sum of Ranks
friendlines	1.00	13	12.81	166.5
	2.00	12	13.21	158.5
	Total	25		

Table 5.5: SPSS output showing the ranks for friendliness.

Test Statistics [b]

	friendlines
Mann-Whitney U	75.50
Wilcoxon W	166.50
Z	-.141
Asymp. Sig. (2-tailed)	.888
Exact Sig. [2*(1-tailed Sig.)]	.894[a]

a. Not corrected for ties.

b. Grouping Variable: VAR00006

Table 5.6: SPSS output showing the results of the Mann Whitney analysis for friendliness.

5.2.1.2 Intelligence

Users were asked to rank the intelligence of the character on a scale of one to seven from unintelligent to intelligent. The normalized scores for intelligence were 0.58 for group 1 and 0.65 for group 2, see Figure 5.17.

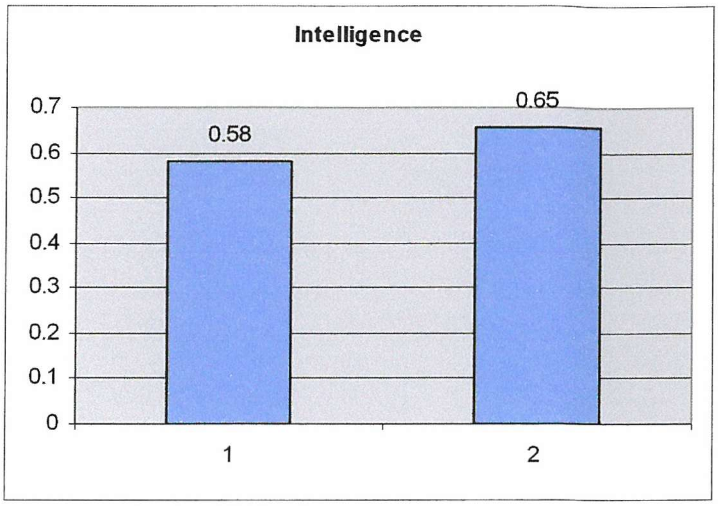


Figure 5.17: The normalized scores for intelligence in the gestures experiment.

Table 5.7 shows N, the number of users, the mean of the rank as explained in and the sum of the ranks (see Section 3.4.2). Table 5.8 shows the results of the Mann Whitney analysis once again showing that the results are not statistically significant; see Section 3.4.2 for further information.

Ranks

	group	N	Mean Rank	Sum of Ranks
intelligence	1.00	13	11.69	152.0
	2.00	12	14.42	173.0
	Total	25		

Table 5.7: SPSS output showing the ranks for intelligence.

Test Statistics [b]

	intelligence
Mann-Whitney U	61.00
Wilcoxon W	152.00
Z	-.958
Asymp. Sig. (2-tailed)	.338
Exact Sig. [2*(1-tailed Sig.)]	.376[a]

a. Not corrected for ties.

b. Grouping Variable: VAR00006

Table 5.8: SPSS output showing the results of the Mann Whitney analysis for intelligence.

5.2.1.3 Pleasantness

Users were asked to rank the pleasantness of the character on a scale of one to seven from unpleasant to pleasant. The normalized scores for pleasantness was 0.62 for both groups, see Figure 5.18.

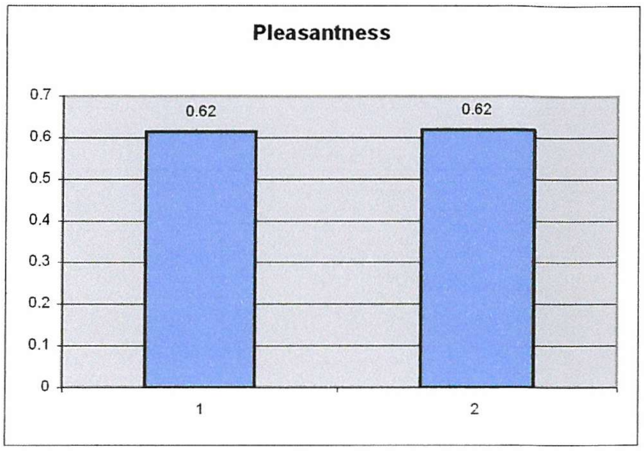


Figure 5.18: The normalized scores for pleasantness in the gestures experiment.

Table 5.9 shows N, the number of users, the mean of the rank as explained in and the sum of the ranks (see Section 3.4.2). Table 5.10 shows the results of the Mann Whitney analysis once again showing that the results are not statistically significant; see Section 3.4.2 for further information.

Ranks

	group	N	Mean Rank	Sum of Ranks
pleasantness	1.00	13	12.81	166.5
	2.00	12	13.21	158.5
	Total	25		

Table 5.9: SPSS output showing the ranks for pleasantness.

Test Statistics [b]

	pleasantness
Mann-Whitney U	75.50
Wilcoxon W	166.50
Z	-.139
Asymp. Sig. (2-tailed)	.889
Exact Sig. [2*(1-tailed Sig.)]	.894 [a]

a. Not corrected for ties.

b. Grouping Variable: VAR00006

Table 5.10: SPSS output showing the results of the Mann Whitney analysis for pleasantness.

5.2.1.4 Interestingness

Users were asked to rank how interesting they found the characters of the character on a scale of one to seven from uninteresting to interesting. The normalized scores for the interesting rating was 0.57 for the first group and 0.51 for the second, see Figure 5.19.

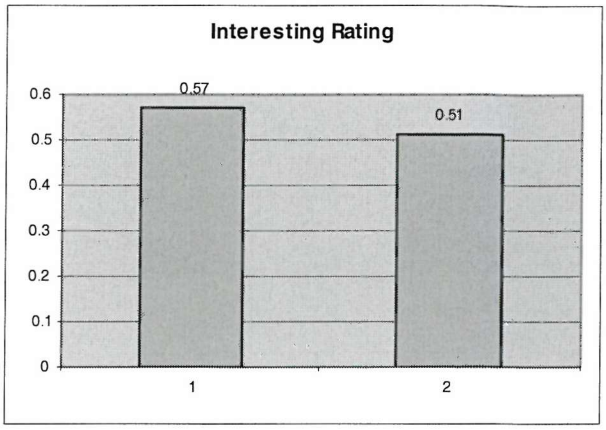


Figure 5.19: The normalized scores for the interesting rating in the gestures experiment.

Table 5.11 shows N, the number of users, the mean of the rank as explained in and the sum of the ranks (see Section 3.4.2). Table 5.12 shows the results of the Mann Whitney analysis once again showing that the results are not statistically significant; see section see Section 3.4.2 for further information.

Ranks

	VAR00006	N	Mean Rank	Sum of Ranks
VAR00003	1.00	13	14.00	182.00
	2.00	12	11.92	143.00
	Total	25		

Table 5.11: SPSS output showing the ranks for how interesting the users thought the character looked.

Test Statistics [b]

	interestingness
Mann-Whitney	65.00
Wilcoxon	143.00
Z	-.725
Asymp. Sig. (2-	.469
Exact Sig. [2*(1-Sig.)]	.503 [a]

a. Not corrected for ties.

b. Grouping Variable: VAR00006

Table 5.12: SPSS output showing the results of the Mann Whitney analysis for how interesting the users thought the character looked.

5.2.1.5 Human like behavior

Users were asked to how much the character behaved like a real human and rate this on a scale of one to seven from "doesn't behave like a human" to "behaves very much like a human". The normalized scores for the human like behaviour rating was 0.55 for the first group and 0.52 for the second see Figure 5.20.

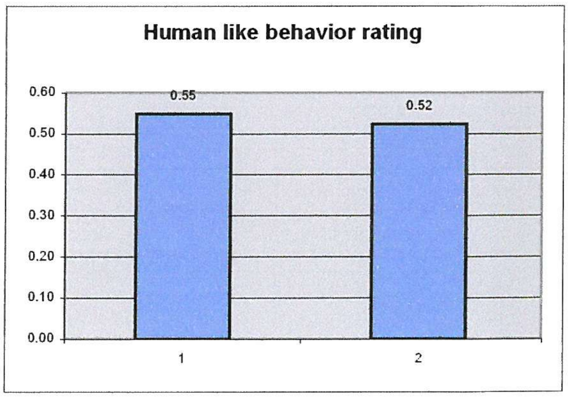


Figure 5.20: The normalized scores for the human like behavior rating in the gestures experiment.

Table 5.13 shows N, the number of users, the mean of the rank as explained in and the sum of the ranks (see Section 3.4.2). Table 5.14 shows the results of the Mann Whitney analysis once again showing that the results are not statistically significant; see Section 3.4.2 for further information.

Ranks

	VAR00006	N	Mean Rank	Sum of Ranks
VAR00003	1.00	13	13.42	174.50
	2.00	12	12.54	150.50
	Total	25		

Table 5.13 SPSS output showing the ranks for how human-like the users thought the character behaved.

Test Statistics[b]

	VAR00003
Mann-Whitney U	72.500
Wilcoxon W	150.500
Z	-.305
Asymp. Sig. (2-tailed)	.760
Exact Sig. [2*(1-tailed Sig.)]	.769[a]

a. Not corrected for ties.

b. Grouping Variable: VAR00006

Table 5.14 SPSS output showing the results of the Mann Whitney analysis for how human-like the users thought the character behaved.

5.2.2 *Information versus presenter*

An unwitting finding was that users do not pay much attention to or even look at the storyteller when they are presented with slides. The number of users who could not see a difference between the two storytellers showed this, in total 15 out of 25 users said they could tell the difference between the two characters, however when asked what the difference was: Three said the voice was different; which cannot be as the presentations were identical but the one without gestures just had the lines of code where gestures are played commented out. One said that the second character sounded "watery and distant". Another said that character A had "better sound". Another suggested that character B had more mouth movements when speaking, which once again cannot be true as it was the exact same character. And finally another test subject suggested that character B seemed friendlier and smiled more, which was not true. Many of these test subjects were appeared surprised when the true difference between the characters was revealed to them.

This means that a total of 7 out of 25 test subjects saw the difference between the characters, one of who commented after the experiment was over that they were consciously trying not to look at he slides and look at the expressions of the character.

One detail worthy of mention is that an attempt was made to obtain a result by only considering data from test subjects who had recognized the difference between the story tellers, however this would not be successful as on the second group only one test subject recognized the differences.

Chi-square (see Section 3.4.1) tests were performed on the first and second groups both including and not including the test subjects who wrongly identified the difference between the storytellers and without and all the test subjects as a whole. The analysis for group including the test subjects who wrongly identified the differences between the storytellers (see Table 5.15) showed that the data was not statistically significant (see Table 5.17). However, for group two (see Table 5.16) it was statistically significant with a degree of confidence of over 95%.

Group 1

	Observed N	Expected N	Residual
.00	7	6.5	.5
1.00	6	6.5	-.5
Total	13		

Table 5.15 SPSS output showing the Observed N and the expected N for group 1.

Group 2

	Observed N	Expected N	Residual
.00	10	6.0	4.0
1.00	2	6.0	-4.0
Total	12		

Table 5.16 SPSS output showing the Observed N and the expected N for group 2.

Test Statistics

	Group 1	Group 2
Chi-Square[a,b]	.077	5.333
df	1	1
Asymp. Sig.	.782	.021

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 6.5.

b. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 6.0.

Table 5.17 SPSS output showing the results of the Chi-squared analysis for both groups.

If the users who wrongly identified the differences between the characters are changed to not having noticed a difference because we considered that the difference they stated was non-existent, then still no statistical significance is found (see Table 5.18 and Table 5.19)

VAR00012

	Observed N	Expected N	Residual
1.00	6	6.5	-.5
2.00	7	6.5	.5
Total	13		

Table 5.18 SPSS output showing the Observed N and the expected N for group 1 with the corrected results.

Test Statistics

	VAR0001
Chi-Square [a]	.077
df	1
Asymp.	.782

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 6.5.

Table 5.19 SPSS output showing the results of the Chi-squared analysis for group 1 with the corrected results.

And finally if we bring all the results together, i.e. group one cancelling the test subjects who stated differences between the characters, which were not there, and group two the results of the chi-square test show that there is a significant difference with a degree of confidence of 95% (see Table 5.20 and Table 5.21).

VAR00010

	Observed N	Expected N	Residual
1.00	7	12.5	-5.5
2.00	18	12.5	5.5
Total	25		

Table 5.20 SPSS output showing the Observed N and the expected N for both groups with the corrected results.

Test Statistics

	VAR0001
Chi-Square [a]	4.840
df	1
Asymp. Sig.	.028

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 12.5.

Table 5.21 SPSS output showing the Observed N and the expected N for both groups with the corrected results.

Overall the results show that users are more interested in the information itself than the character presenting it, which on hindsight isn't a surprising result. It is inconclusive why the users in group one seemed more able to recognize the differences between the characters than group two.

5.3 Discussion and Conclusion

The key finding in this section is that the information being presented is more important that the presenter and users seem to pay little attention to the character presenting some of this information even when the character is presenting the information in verbal form, like the storytelling setting here. This brings up two important questions; can the anthropomorphic character presenting the information enhance communication of information in certain situations? And is the anthropomorphic character presenting the information at all necessary? It is not surprising that we find ourselves asking that question at some point during the course of these studies. This is a contentious issue in the field; Lanier professes against the "evils" of autonomous agents (Lanier 1995; Lanier 1996), Walker has shown negative results with anthropomorphic interfaces when compared to text only interfaces (Walker, Sproull et al. 1994), Shneiderman proposes that user anthropomorphic agents are largely unnecessary as there are other methods to achieve the advantages agents are said to provide. On the other hand there are also many advocates of anthropomorphic interfaces, during a panel discussion at CHI-92 (Don, Brennan et al. 1992) Laurel argued for anthropomorphism, Koda claims that at least for entertainment software anthropomorphic agents are useful. These issues are discussed in depth in Chapter 7.

6 Emotion

This chapter reports on experiments investigating how anthropomorphic characters displaying accurate emotions according to the OCC model (Ortony, Clore et al. 1988) affect user's opinions and acceptance of anthropomorphic characters.

We find that users show a clear preference for the character that display emotions as compared to character that does not. T0his is backed up by user comments.

The character displaying emotion is rated as being friendlier, more intelligent, more pleasant and more interesting to look at.

6.1 Introduction

We considered two possible scenarios for this experiment; one was an interactive Web-searching assistant like that in Section 4.2.3 but this time with an emotional component or a one-way-interaction football commentator. The football commentator experiment was considered most appropriate for this situation, this decision was reached considering two factors: our findings in Chapter 4; that lower levels of interaction replicate results of higher levels of interactions and that we are used to seeing highly emotional and exaggerated displays of emotion from football commentators. There was concern that highly emotional character may seem out of place in a web-searching assistant and that more subtle emotional displays might be missed by the user as occurred with gestures in Chapter 5.

6.2 Initial Considerations

A biased football commentator was chosen as the setting for the experiment and since the experiment took place at the University of Southampton it was decided to have commentator that was biased towards the Southampton Football Club (Saints). We considered that a biased commentator would be better at displaying clear emotions than an impartial one. A commentator that supports a particular team could possibly get angry when refereeing decisions go against the its team, get exited when the team is about to score, get scared when one of the team's players appears to be injured, happy when the team scores, surprised when the opposition shoots for the goal, etc.

The emotions displayed by our football commentator (see Figure 6.1) are based on Ekman's universal expressions (Ekman and Friesen 1975); happy, sad, anger, fear, disgust and surprise. We excluded disgust from our experiment as a small pilot experiment indicated our chosen expression for disgust wasn't easily identifiable and most users identified it by a process of elimination.

Figure 6.1: Emotions (happy, sad, fear, angry and surprise) displayed by the football commentator.

Given our finding in Chapter 5 that users tend to look at the information being presented and pay little to no attention to the presenter, it was decided that users would be given very little to look at apart for the commentator. The whole of the football commentary takes place on a nearly blank web-browser window, the commentator takes up a central position on the screen and the only other thing to look at is a title "Southampton vs. Ipswich", which does not change thought the whole virtual football match (see Figure 6.2).

Figure 6.2: Football commentator in action.

6.3 Emotional Model

The emotional model for the commentator is based on the OCC model of emotion (Ortony, Clore et al. 1988). The emotional states (see Figure 6.3) of the commentator are established based on the circumstances of the match it is commenting on.

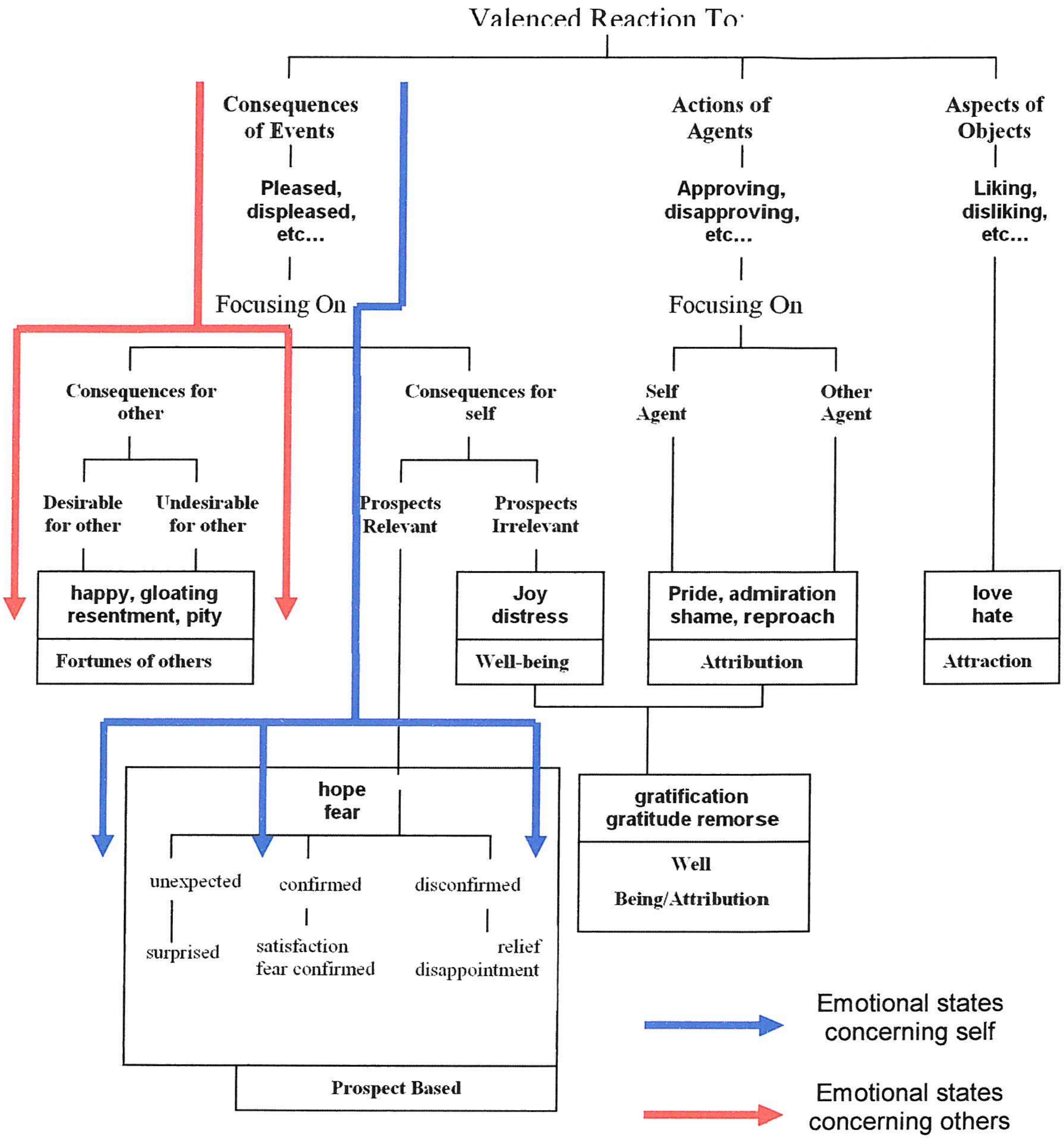


Figure 6.3: OCC model of emotion showing emotional states exhibited by football commentator.

As mentioned in section 6.2, the facial expressions of the commentator are based on Ekman's universal expressions, which are matched to the emotional states in Figure 6.3. There are two distinct trails starting from consequences of events, one going down the consequences for self route and another down the consequences for others route. The first path concerns scoring possibilities for either team and the second concerns fouls and refereeing decisions and fouls.

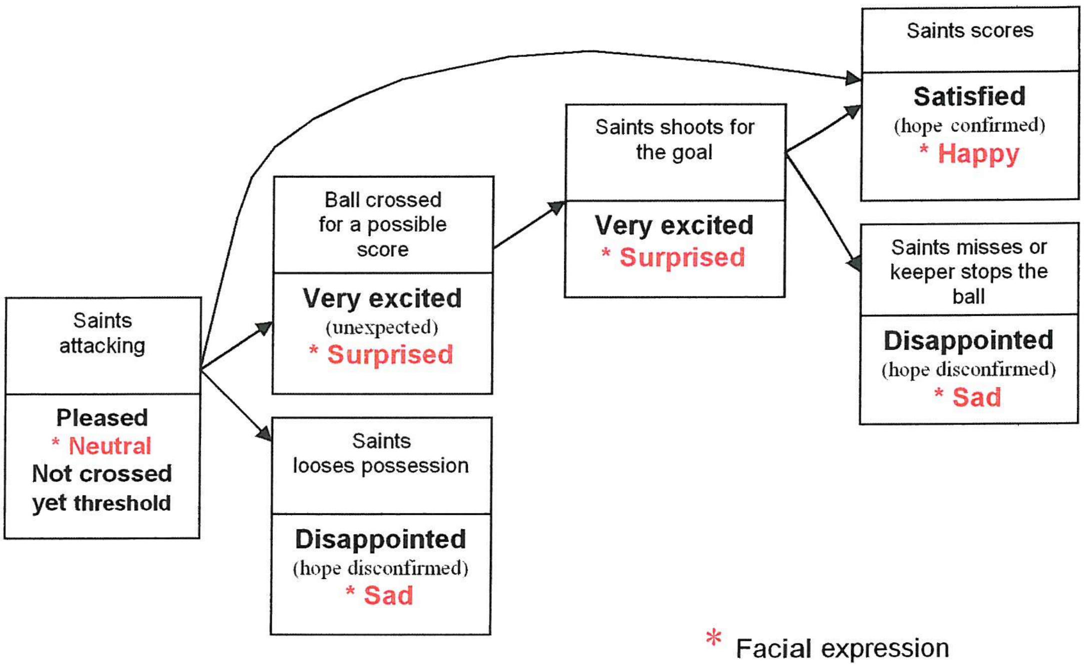


Figure 6.4: Emotional states concerning a Saints attack.

Figure 6.4 shows the emotion types and matching facial expressions for one of the consequences for self trails concerning a Saints attack. This diagram can be seen as two smaller separate trees; one starting from "Saints attacking" with the children "Saints scores", "Saints looses possession" and "Saints scores", the beginning with "Saints shoots for the goal" with the children "Saints scores" and "Saints misses or keeper stops the ball". On the first tree we see that the commentator is pleased yet the expression is neutral, this is taking into account the assumption that there may exist a threshold that needs to be crossed before a person exhibits a facial expression (Ekman 1999).

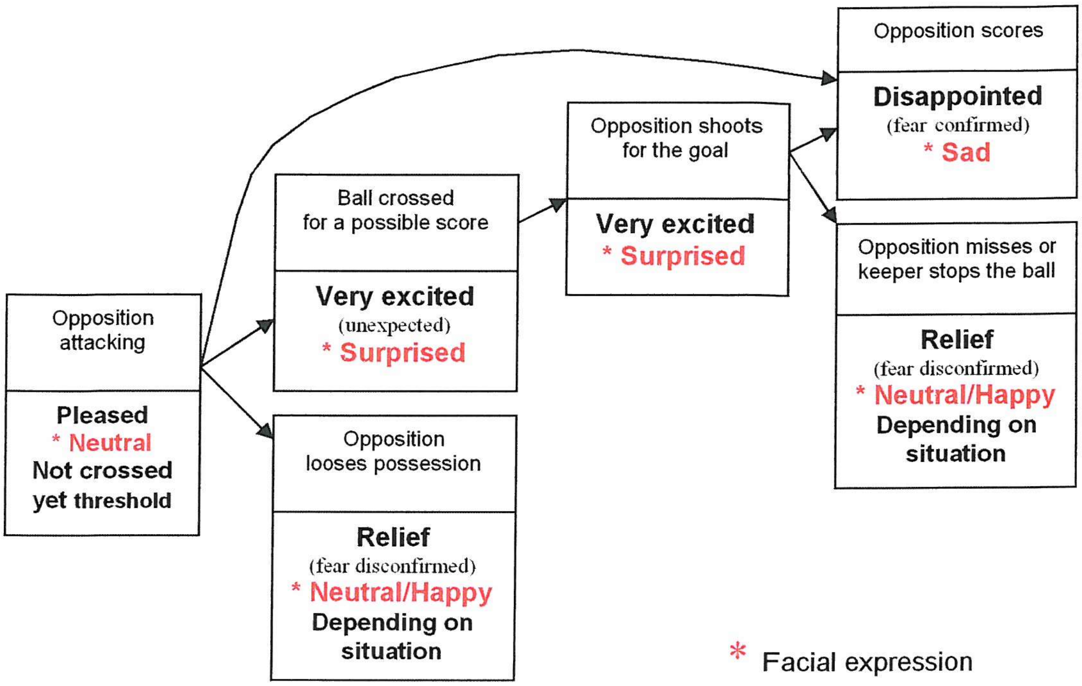


Figure 6.5: Emotional states concerning an Opposition attack.

Figure 6.5, which concerns an opposition attack, is very similar to Figure 6.4. We should mention that when the commentator experiences relief as opposition looses possession and opposition misses the goal, it was thought that even though the commentator is biased having it look overtly pleased every time the opposition looses the ball or misses a goal might prompt the test subjects into thinking the commentator is arrogant.

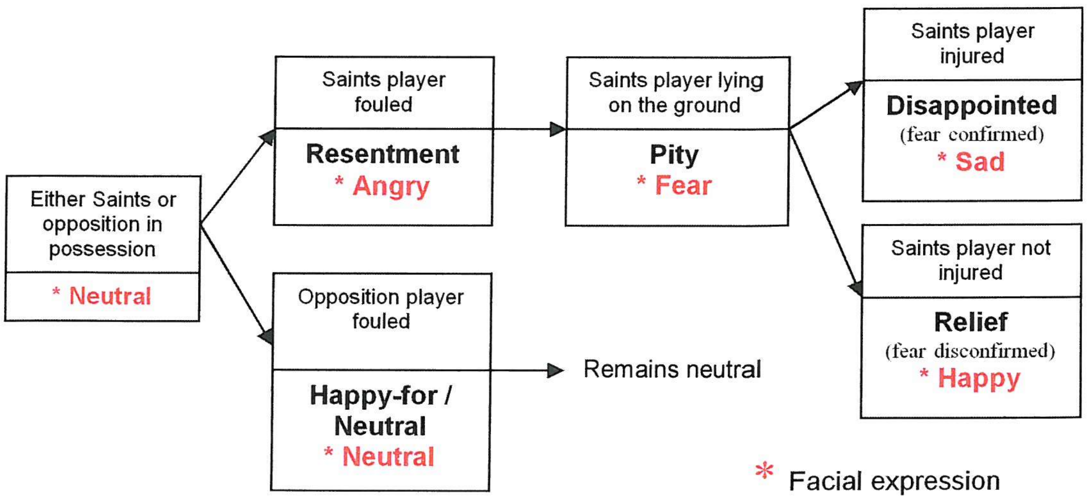


Figure 6.6: Emotional states concerning fouls and injuries.

In Figure 6.6 we can see the emotional states concerning fouls and injuries, this shows emotion types and matching facial expressions for two of the consequences for others trails and one of the consequences for self. It is important to note than when an opposition player is fouled the commentator remains neutral, yet, if the commentator is happy for whoever committed the foul it could conceivably gloat as indicated by the OCC model (see Figure 6.3). However, once again, it was felt that a commentator that gloats when fouls are committed would seem arrogant. Clearly it would also be possible for the commentator to feel angry at his own team when a particularly bad foul is committed against an opposition player, however, we won't be creating this situation during this experiment. As for opposition injuries, once again the commentator could feasibly gloat or feel happy for it's own team or pity for the opposition player, however for this experiment the commentator will always remain neutral.

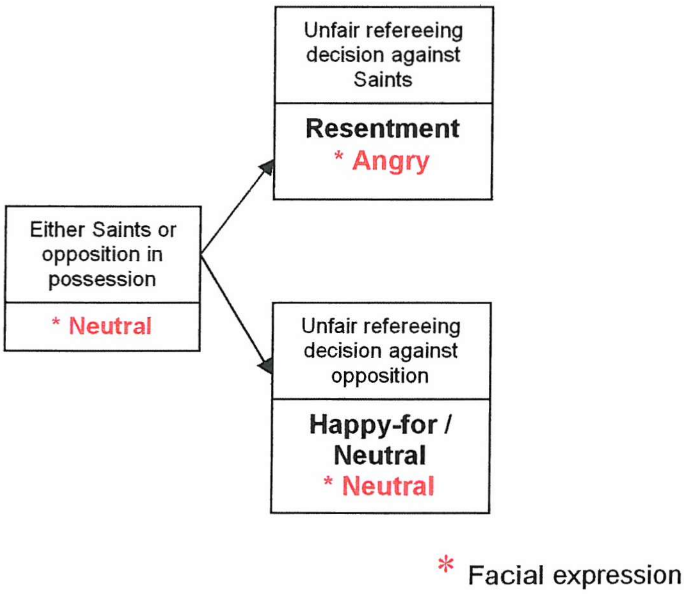


Figure 6.7: Emotional states concerning unfair refereeing decisions.

Figure 6.7 shows the emotional states concerning unfair refereeing decisions, showing emotion types and matching facial expressions for one of the consequences for others trails. Once again to make sure the commentator is not seen as arrogant it doesn't gloat when unfair refereeing decisions against the opposition take place.

The emotional model expressed by Figure 6.4-Figure 6.7 can be simplified into a much less complex event-expression model, see Table 6.1, here every event which could induce an emotional state from the commentator is matched to one of Ekman's universal facial expressions.

Event	Facial Expression
Ball crossed by Saints for a possible score	Surprised
Saints looses possession	Sad
Saints shoots for the goal	Surprised
Saints scores	Happy
Saints misses or keeper stops the ball	Sad
Ball crossed by opposition for possible score	Surprised
Opposition looses possession	Neutral / Happy
Opposition shoots for the goal	Surprised
Opposition scores	Sad
Opposition misses or keeper stops the ball	Neutral / Happy
Saints player fouled	Angry
Opposition player fouled	Neutral / Happy / Angry (Depending on situation, but for this experiment the commentator will always remain neutral)
Saints player lying on the ground	Fear
Saints player injured	Sad
Saints player not injured	Happy
Unfair refereeing decision against Saints	Angry
Unfair refereeing decision against opposition	Neutral / Happy / Angry (Depending on situation, but for this experiment the commentator will always remain neutral)
Opposition player injured	Neutral / Happy / Sad (Depending on situation, but for this experiment the commentator will always remain neutral)
Opposition player not injured	Neutral / Happy / Sad (Depending on situation, but for this experiment the commentator will always remain neutral)

Table 6.1: Simplified emotional model.

6.4 Experimental design

This section describes the experimental design in detail

6.4.1 The Commentator

The commentator's range of facial expressions is based on Ekman and Friesen's universal facial expressions ideas (see Section 6.3). These expressions are: happy, sad, surprise, fear and anger. As we didn't want to run the risk of having facial expressions that were ambiguous, given that expressions in simple abstract characters are easier to recognize (McCloud 1994), it was decided to use the smiley face character form Chapter 4 as our commentator.

The voice used is Microsoft Male 1. The only emotion in the voice is the exclamation marks used on the match script that the text to speech software is able to interpret, it is worth pointing out that exclamation marks were used in both of the experimental groups.

6.4.1.1 The expressions

In this section the characteristics of the universal facial expressions as defined by Ekman and Friesen (Ekman and Friesen 1975) are described.

Happiness: Raising and lowering of mouth corners characterize happiness. The frames of animation used for happiness can be seen in Figure 6.8.

Figure 6.8: Frames used for the happy facial expression.

Sadness: Lowering of mouth corners raise inner portion of brows characterize sadness. The frames of animation used for sadness can be seen in Figure 6.9.

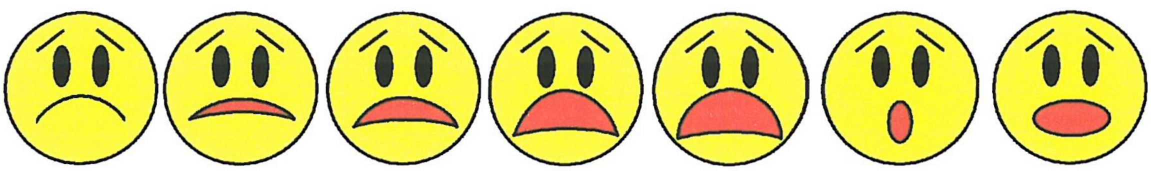


Figure 6.9: Frames used for the sad facial expression.

Surprise: Eyes opening wide to expose more white and jaw dropping slightly are characteristics of the surprised expression. The frames of animation used for surprise can be seen in Figure 6.10.

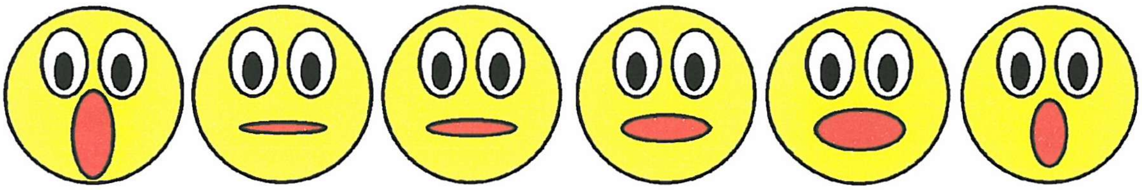


Figure 6.10: Frames used for the surprised facial expression

Fear: Brows raised, eyes open wide and mouth opening slightly are characteristics of the fear expression. The frames of animations used for fear can be seen Figure 6.11.

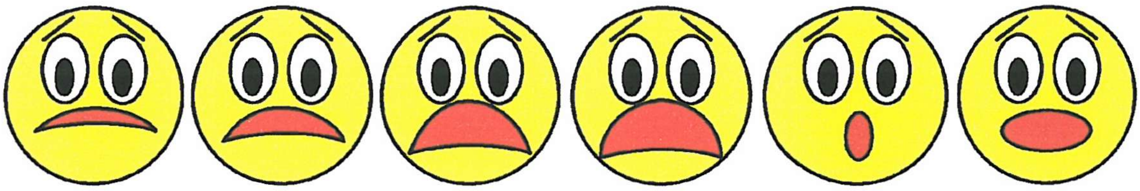


Figure 6.11: Frames used for the fear facial expression

Anger: Brows lowered, lips pressed firmly and eyes bulging are all characteristics of the universal expression for anger. The frames of animations used for this expression can be seen Figure 6.12.

Figure 6.12: Frames of animation used for the anger expression

6.4.2 Experiment structure

The whole experiment took part in a single sitting, there were two user groups one group evaluated a commentator capable of displaying the facial expressions described above and the other, a control group, evaluated a commentator with no facial expressions. The questions and the football commentator were all displayed on a browser window.

The experiment was divided into several stages; the first stage was designed to ensure that test subjects recognized the facial expressions exhibited by the commentator. Users were shown pictures of the commentator displaying emotions and they were asked to match these to a list provided, see Figure 6.13.

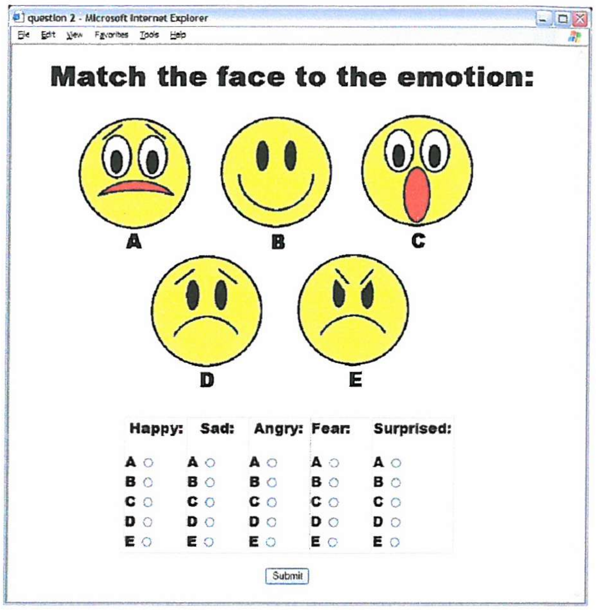


Figure 6.13: Question 1 for emotion experiment.

During the second stage of the experiment (see Figure 6.2) users were shown the football commentator, group one was shown the commentator capable of facial expressions and the control group was shown the character incapable of displaying facial expressions. It should be noted that both scripts, Microsoft Agent programmed with VBScript, are exactly the same but the script for the control group has had the facial expression lines deleted, see Appendix 7.

Then users were asked to assess the character (see Figure 6.14), the attributes examined for this experiment were: friendliness, intelligence, pleasantness and how interesting they are. For information on how these attributes were decided upon refer to page 52.

Users were then informed that the experiment was divided into two groups and in which group they were (see Figure 6.15). This page links to the other group's commentator, so when the users follow the link they are presented with the commentator that the other group saw.

UNIVERSITY OF SOUTHAMPTON LIBRARY

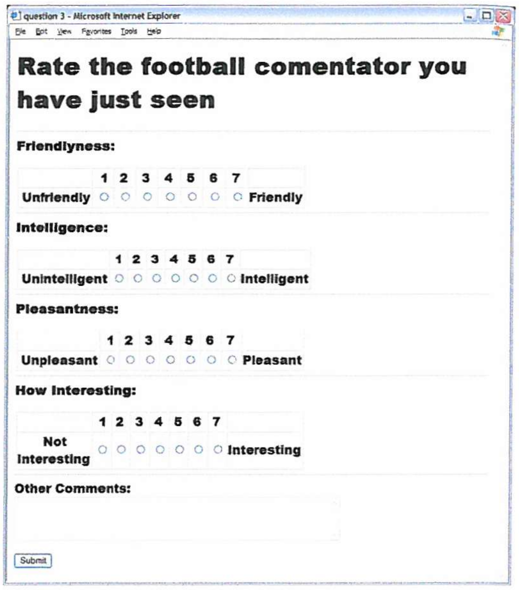


Figure 6.14: Question 2-3 for emotion experiment.

Finally when the commentator finishes users are asked which commentator they preferred and why, see Figure 6.16.

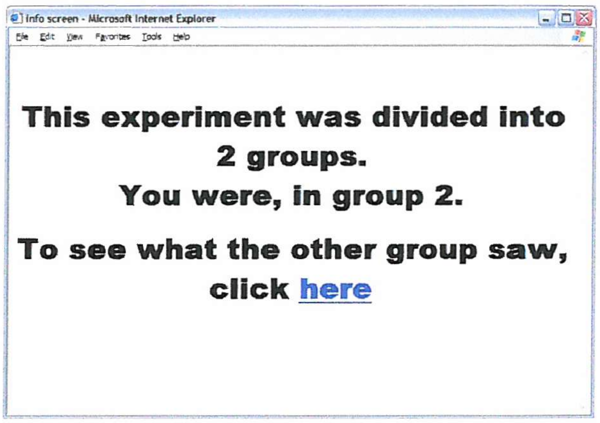


Figure 6.15: Screen leading up to final commentator.

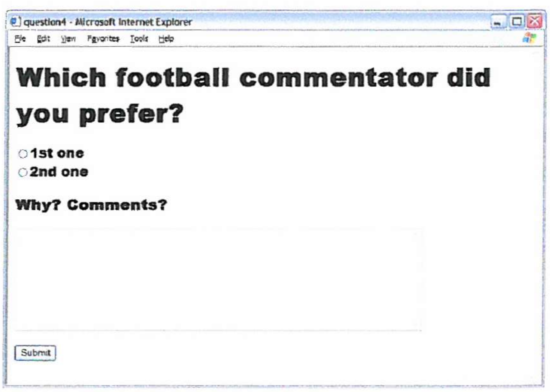


Figure 6.16: Final commentator choice question.

6.5 Results and analysis

This section presents the results for the emotion experiment. The results are divided in subsections corresponding to emotion recognition, the attributes examined and character selection. As mentioned is section 6.4.2 the users were divided into two groups, see Table 6.2.

Population information	Test subjects were obtained from within the Electronics and Computer Science department at the University of Southampton.
Number of subjects in group 1	8
Number of subjects in group 2	7

Table 6.2: Test subject profile for emotion experiment.

To analyse the results we gave the scales on the questionnaires (see Figure 6.14) a weighting. The scales were weighted for 1 to 7. The scores were totalled up and normalized from 0 to 1. The results for each attribute examined were analysed using the non-parametric Mann-Whitney test in order to evaluate if the differences observed between the two groups are statistically significant.

6.5.1 Emotion Recognition

Test subjects were asked to match faces to a list of emotions as seen in section 6.4.2. Twelve out of the fifteen test subjects recognized all of the emotions correctly. The three test subjects who failed to correctly identify emotions, confused fear and sadness.

The frequencies of these choices were recorded and analysis of these was carried out. To compare the frequencies of each group, we state the null hypothesis that the user gave the same frequency to each of the emotions, i.e. 1/5. The alternative hypothesis is that the character was matched to the correct emotion.

The chi-square test was employed to determine the level of significance of the identified emotions. These tests showed that all emotions were matched correctly with a confidence rating greater than 99%.

6.5.2 Friendliness

Test subjects were asked to rank the friendliness of the football commentators using a scale of one to seven corresponding to unfriendly to friendly respectively. These scores were normalized and are presented in Figure 6.17. Test subjects in the group exposed to the commentator capable of displaying facial expressions (group 1) scored the commentator as being friendlier with a score of 0.80 compared to 0.65 in the group exposed to the commentator incapable of displaying facial expressions (group 2).

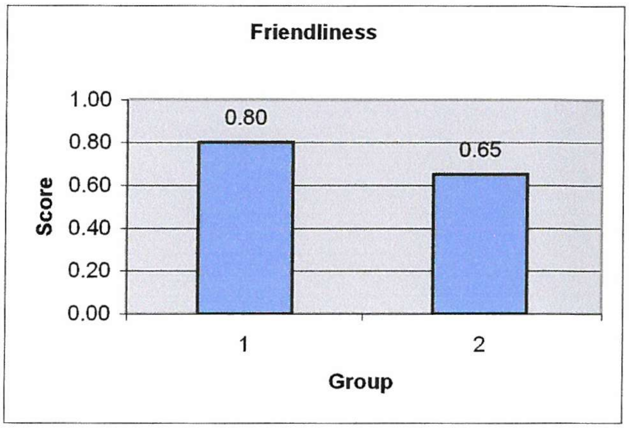


Figure 6.17: Normalized scores for friendliness in the emotion experiment.

Ranks

	group	N	Mean Rank	Sum of Ranks
friendliness	1.00	8	9.88	79.00
	2.00	7	5.86	41.00
	Total	15		

Table 6.3: SPSS output showing the ranks for friendliness.

Table 6.3 shows N (number of test subjects), the mean of the rank and the sum of the ranks as explained in Section 3.4.2. Table 6.4 shows the results of the Mann-Whitney analysis, for this case the results are weakly significant with a confidence greater than 90%.

Test Statistics

	friendliness
Mann-Whitney U	13.000
Wilcoxon W	41.000
Z	-1.905
Asymp. Sig. (2-tailed)	.057
Exact Sig. [2*(1-tailed Sig.)]	.094

a Not corrected for ties.
b Grouping Variable: VAR00002

Table 6.4: SPSS output showing the results of the Mann Whitney analysis for friendliness.

6.5.3 *Intelligence*

Test subjects in the group exposed to the commentator capable of displaying facial expressions (group 1) scored the commentator as being more intelligent with a score of 0.70 compared to 0.53 in the group exposed to the commentator incapable of displaying facial expressions (group 2), see Figure 6.18.

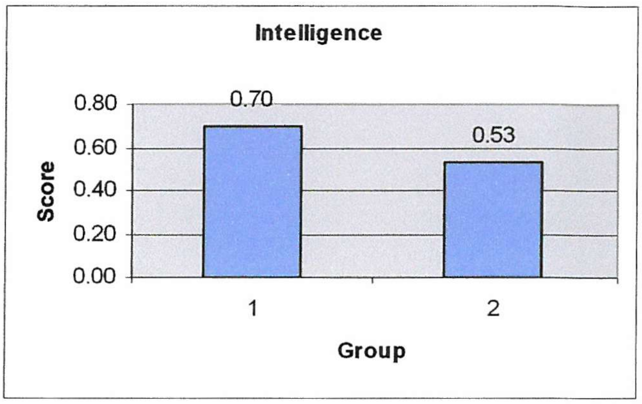


Figure 6.18: Normalized scores for intelligence in the emotion experiment.

Table 6.5 shows N (number of test subjects), the mean of the rank and the sum of the ranks as explained in Section 3.4.2. Table 6.6 shows the results of the Mann-Whitney analysis, for this case the results are strongly significant with a confidence greater than 95%.

Ranks

	group	N	Mean Rank	Sum of Ranks
intelligence	1.00	8	10.13	81.00
	2.00	7	5.57	39.00
	Total	15		

Table 6.5: SPSS output showing the ranks for intelligence.

Test Statistics

	intelligence
Mann-Whitney U	11.000
Wilcoxon W	39.000
Z	-2.044
Asymp. Sig. (2-tailed)	.041
Exact Sig. [2*(1-tailed Sig.)]	.054

a Not corrected for ties.
b Grouping Variable: VAR00002

Table 6.6: SPSS output showing the results of the Mann Whitney analysis for intelligence.

6.5.4 *Pleasantness*

Test subjects in the group exposed to the commentator capable of displaying facial expressions (group 1) scored the commentator as being more pleasant with a score of 0.77 compared to 0.59 in the group exposed to the commentator incapable of displaying facial expressions (group 2), see Figure 6.19.

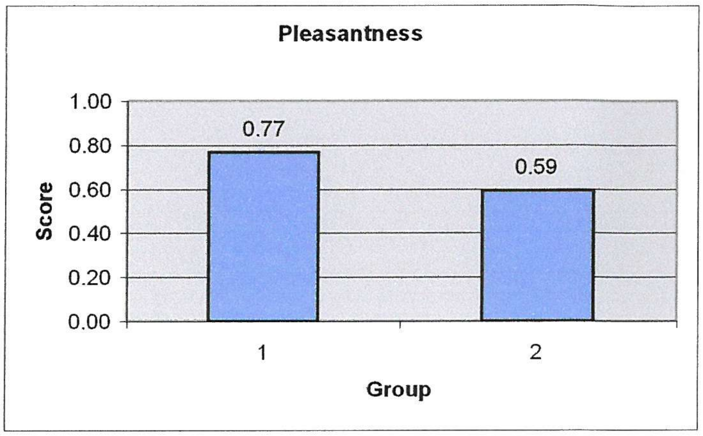


Figure 6.19: Normalized scores for pleasantness in the emotion experiment.

Table 6.7 shows N (number of test subjects), the mean of the rank and the sum of the ranks as explained in Section 3.4.2. Table 6.8 shows the results of the Mann-Whitney analysis, for this case the results are weakly significant with a confidence greater than 90%.

Ranks

	group	N	Mean Rank	Sum of Ranks
pleasantness	1.00	8	10.00	80.00
	2.00	7	5.71	40.00
	Total	15		

Table 6.7: SPSS output showing the ranks for intelligence.

Test Statistics

	pleasantness
Mann-Whitney U	12.000
Wilcoxon W	40.000
Z	-1.900
Asymp. Sig. (2-tailed)	.057
Exact Sig. [2*(1-tailed Sig.)]	.072

a Not corrected for ties.
b Grouping Variable: VAR00002

Table 6.8: SPSS output showing the results of the Mann Whitney analysis for pleasantness.

6.5.5 *How interesting*

Test subjects in the group exposed to the commentator capable of displaying facial expressions (group 1) scored the commentator as being more interesting with a score of 0.77 compared to 0.41 in the group exposed to the commentator incapable of displaying facial expressions (group 2), see Figure 6.20.

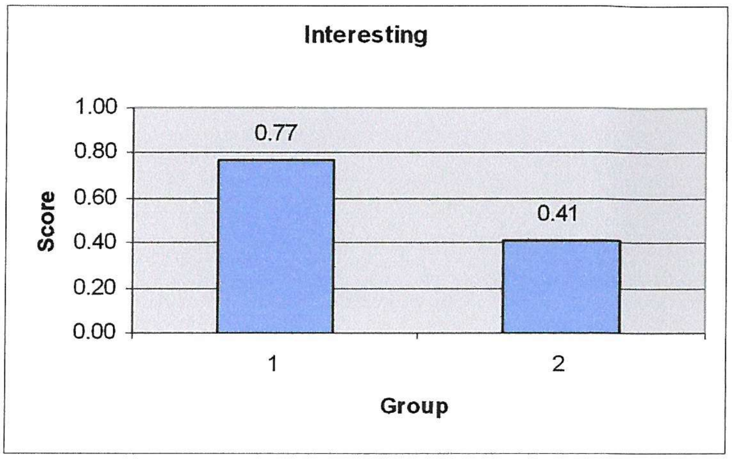


Figure 6.20: Normalized scores for the level of interest in the emotion experiment.

Table 6.9 shows N (number of test subjects), the mean of the rank and the sum of the ranks as explained in Section 3.4.2. Table 6.10 shows the results of the Mann-Whitney analysis, for this case the results are strongly significant with a confidence greater than 95%.

Ranks

	group	N	Mean Rank	Sum of Ranks
interest	1.00	8	11.31	90.50
	2.00	7	4.21	29.50
	Total	15		

Table 6.9: Normalized scores for the level interest in the emotion experiment.

Test Statistics

	interest
Mann-Whitney U	1.500
Wilcoxon W	29.500
Z	-3.132
Asymp. Sig. (2-tailed)	.002
Exact Sig. [2*(1-tailed Sig.)]	.001

a Not corrected for ties.

b Grouping Variable: VAR00002

Table 6.10: SPSS output showing the results of the Mann Whitney analysis for level of interest.

6.5.6 Commentator choice

All of the test subjects chose the commentator that displayed facial expressions as their favourite character. Considering previous results that the commentator displaying facial expressions was rated as being friendlier, more intelligent, more pleasant and more interesting to look at this is not a surprising result.

We get a further insight into why users favoured the commentator displaying facial emotions from their feedback. The mention of more useful information being provided by the facial expressions occurs in a number of occasions:

- "Facial expressions helped explain the mood."
- "Expressions underline the action."
- "The emotions expressed by the character helped me follow the game."
- "It was much easier to understand the emotion the commentator was attempting to convey, partly because the voice wasn't quite accurate enough to convey the emotion alone, so the character's facial expression helped. It also helped to keep one's attention for longer."
- "The variety is interesting although it doesn't convey any more information, it does provide different information."
- "The expression would also contribute to the meaning of a speech."
- "The expressions do make up for the relative lack of emotion in the voice of the commentator and also serve as a cue when you want to pay more attention in the eventful parts of the commentary."
- "Well, it is easier to know who is winning even if you are not watching the match."
- "The facial expressions made it seem like the commentator cared more about what was happening."
- "It illustrated the critical portions of the game like which parts were exciting, boring, etc."

Most test subjects mentioned that facial expressions provided extra information that proved useful. The author proposes that accurate display of emotions by an anthropomorphic agent can enhance or provide extra information to the user and it is a key factor to be considered by anthropomorphic interface designers.

6.6 Summary

We have seen how test subjects perceive anthropomorphic characters that display accurate emotions as being friendlier, more intelligent, more pleasant and more interesting to look at and how test subjects overwhelmed prefer anthropomorphic agents which display accurate emotions.

We also propose that accurate display of emotions provides extra or enhances information conveyed by the anthropomorphic agent and we propose that this is a key factor to be consider in anthropomorphic interface design.

The author considers that it is highly likely that the reason the commentator displaying facial expressions was so readily accepted by users was because it displayed a believable level of emotion, it is common for football commentators to be highly emotional in the real world, therefore test subjects found it easy to accept the highly emotional facial expressions used for this experiment.

7 Conclusions, Discussion and Future Work

This chapter presents conclusions reached during our research, discusses issues building on the results we have discovered as part of this thesis and presents possible future work.

7.1 Conclusions

In this section we list findings reached during the course of this research and briefly discuss these. This section is divided into subsections that categorize the findings.

7.1.1 Levels of Anthropomorphism

The realistic faces were seen as less friendly and even scary: We propose the *scary* factor is due to the character's behaviour not matching the user's expectations of the human looking face. We put forward that this is not a general finding for and that it can be seen as a technological barrier that will be surmounted with time.

The more abstract a character the more friendly/pleasant it seems: We consider the possibility that less realistic more abstract characters leave more open to the imagination and thus are less likely to be disliked.

The more abstract the character the less intelligence users will attribute to it: This was an expected finding, we consider that as intelligence is thought as being specifically a human characteristic it is understandable that users see the more human like characters as more intelligent.

The more abstract the character the less interesting it is to look at: It is not surprising that users rate the simpler characters as less interesting to look at than more complex ones as there is less detail to see.

Although partial interaction does lower the scores attributed by the users it does not appear to change the overall patterns displayed. Therefore when developing future studies the use of partial interaction could prove a useful tool in developing quicker experiments.

7.1.2 Voice

Subjects were not always aware that different voices were employed during the experiment. This may be because subjects were focusing on other aspects of the interaction instead of the particular voice used. It is also worth mentioning that voice 2 scored lowest in understanding. It is thought this may be due to the voice skipping, as the laptop being used for the experiment sometimes could not cope with that particular TTS engine. It appears that tailoring the software to the hardware being used is the key to the success of the interface and the particular sound of the TTS being used might be relatively unimportant.

7.1.3 Character Choice

The users favoured a moderately abstract character: A cartoon character was favoured over a smiley face and a realistic face. We propose that its success is because it scores well in all characteristics as opposed to the other characters, which score very high for some features and very low for others. However we do not propose this is a general finding, perhaps for different applications with different capabilities might produce different results, we consider if very likely that different applications are suited to different looking characters.

We present a method of calculating how much like a human the user is treating the agent and experimental results indicate that users treat their favourite characters more like humans than characters they do not like as much. This finding could be exploited to build a system which analyses user and agent interaction to determine anthropomorphic agent acceptance without carrying out user experiments.

7.1.4 Information and Presenter

We show that the information being presented is more important than the presenter and users seem to pay little attention to the character presenting some of this information even when the character presented the information in verbal form. This is an issue, which should be considered carefully when designing a conversational interface.

7.1.5 Emotions

We showed how test subjects perceive anthropomorphic characters that display accurate emotions as being friendlier, more intelligent, more pleasant and more interesting to look at and how test subjects overwhelmed prefer anthropomorphic agents which display accurate emotions.

We also propose that accurate display of emotions provides extra or enhances information conveyed by the anthropomorphic agent and we propose that this is a key factor to be consider in anthropomorphic interface design.

We highlight that it is highly likely that the reason the commentator displaying facial expressions was so readily accepted by users was because it displayed a believable level of emotion, it is common for football commentators to be highly emotional in the real world, therefore test subjects found it easy to accept the highly emotional facial expressions used for this experiment.

In this chapter we discuss concerns expressed by the research community on anthropomorphic agents and set out future work to help answer some of these concerns.

7.1.6 Guidelines

In this section we summarize our most important findings and present them as guidelines for creator of anthropomorphic interfaces to follow:

1) **Carefully match the level of anthropomorphic abstraction of the character to the technology at your disposal.** Our studies have shown that users expect more realistic characters to behave more like humans, in order to fulfill user expectations the look of the character needs to be closely matched to the technology available.

2) **Consider when and how to present information.** We have shown how the information being presented is more important to the user than the presenter and users seem to pay little attention to the character presenting some of this information even when the character presented the information in verbal form. This is an issue, which should be considered carefully when designing a conversational interface; the timing for presenting information as well as the actual need for preserving the anthropomorphic character on screen when presenting information both need to be considered.

3) **Model emotions accurately.** We have showed how characters that display accurate emotions as being friendlier, more intelligent, more pleasant and more interesting to look at. We also theorize that accurate display of emotions provides extra or enhances information conveyed by an anthropomorphic agent.

4) **Tailor the look of the character to your application.** We found that the look of the character can greatly affect user expectations of it. We propose that when designing a character its look should be tailored to the specific application in mind. For example we have seen how more abstract characters are seen as being friendlier, so perhaps if we were designing a storytelling software aimed at young children then a more abstract character would be more suitable than a realistic character.

7.1.7 Problems with our experiments

This section details aspects of our experiment that on hindsight may have been carried out differently.

- It would have been useful to carry out a small experiment to make sure that users were correctly interpreting the attributes we were testing for.
- During our emotions experiments it would have been useful to compare it to a third group. Our finding showed that emotions enhanced the information being presented, but it would be interesting to compare it with a third group where users are just presented with a voice and a background that flashes whenever important events happen during the game.
- It would have been interesting to compare male versus female characters during the first level of abstraction experiments.

We maintain that these issues do not alter our conclusions, but should be kept in mind for future expansion of this research.

7.2 Discussion

7.2.1 Do we really need anthropomorphic agents?

We posed the questions "when are anthropomorphic interfaces necessary?" and "are the necessary at all?" in section 5.3. In this section we discuss these two questions. Anthropomorphic agents have for long been a dream in both science fiction and real science, here we see what both science fiction and science has to say about the need, dangers, possibilities, etc. of anthropomorphic interfaces.

7.2.2 Science Fiction

Since anthropomorphic user interfaces are relatively new and most, if not all, of the anthropomorphic interfaces available at the moment are rudimentary at best, it was decided to look at science fiction to try and get an insight into the human psyche concerning expectations and acceptance of seemingly intelligent, anthropomorphic interfaces.

At one level it appears to be what people want. Since the beginning of the machine age humans have dreamed of intelligent machines as found in Isaac Asimov works of science fiction. He views the idea of anthropomorphic robots as useful tools to humans, he first introduces this idea in his book Strange Playfellow or Robbie as it later became known. Since then we have seen many friendly/useful anthropomorphic robots and interfaces in science fiction, for instance Max Headroom, Data for Star Trek: Next Generation, No 5 from short circuit, KITT from Knight Rider, etc...

Conversely, before Isaac Asimov introduced the idea of robots being useful tools the prevailing idea of robots in science fiction was as "menaces or sort of wistful little creatures" that generally turned on their creators to punish them for playing god and creating life. This view is still prevalent today and is still seen in many modern science fictions works, for instance HAL from Space 2001: Space Odyssey murders the crew, The Terminator is a highly anthropomorphic robot used to exterminate the human race and even robots which are predominantly portrayed as being friendly and useful do frequently brake down and end up causing havoc and putting humans into danger.

This appears to highlight the fear or apprehension that people have of intelligent machines and even when humans have accepted anthropomorphic computers science fiction stresses that there is an underlying danger of seemingly intelligent machines.

7.2.3 Examples and questions

It is often quoted stated people tend to anthropomorphise computers (Reeves and Nass 1996) and hence a natural step is to anthropomorphise the interface. The author does not believe this is such an obvious and direct step, if science fiction stories on computers with anthropomorphized interfaces are based on genuine human fears people do show resistance toward anthropomorphic interfaces and thorough research on how people accept agents with intelligent approaching that of a human. For instance: Could users develop feelings for anthropomorphic interface agents once these agents behave close enough to humans? To a certain extent we have already witnessed one case where people have become highly attached to anthropomorphic characters, Tamagotchis (see Figure 7.1). Tamagotchis are small virtual pets that inhabit egg shaped key rings, they were hugely popular in the late 1990s. The Tamagotchi character is a fictional creature that grows and develops in different ways according to how the owner takes care of it, or plays with it or disciplines it, etc. The onscreen representation of this creature is rather crude, the resolution of the display is 32x16 pixels and it is monochromatic. Nevertheless, despite this crudeness, there were reports of people becoming highly attached to these virtual pets. Tamagotchis were banned from many British schools as children were unable to stop interacting with them even during lessons and perhaps even more bizarrely there were and still are several virtual graveyards and cemeteries for Tamagotchis, there was even one real cemetery in Germany. Maldonado has reported that children deeply care for virtual pets using Tigrito (Maldonado, Picard et al. 1998), an effective interactive character designed to study children's sense of engagement and relationship with virtual toys and in different modes of interaction. If this effect were to occur with anthropomorphic agents for non entertainment applications, what would this mean to the way humans and computers interact if the users begging to get attached to their interface agents? The author can envisage great resistance from the less computer inclined members of the public when they begin to notice this effect for themselves.

Figure 7.1: Tamagotchi toys.

Science fiction also often explores the opposite effect, that of artificial humans developing feelings for real humans and even human and robots developing feeling for each other an example can be found in *Blade Runner*, a 1982 film based on novel *Do androids dream of electric sheep?* by Philip K. Dick (Dick 1968). In Blade Runner a human character Deckard and "replicant" Rachel fall in love and leave together even though the relationship is not acceptable to other people or compatible with Deckard's job and lifestyle.

The development of real emotions by robots and anthropomorphic computers is often portrayed as the solution or ultimate goal to solving many of the dangers of intelligent machines. It is often believed that the main difference between humans and intelligent machines is the ability to experience emotions. On the other hand the instability of emotions is also portrayed as a danger for machines, which are supposed to be infallible. How would human users react to a computer that appears to genuinely care for them?

7.2.4 Acceptance by the general public

Both in science fiction and the real world some anthropomorphic interfaces seem to be more acceptable that others. For instance many science fiction robots seem very cold and often don't seem very welcoming to humans. In real life characters like Ananova are readily accessible and very popular yet others seem annoying and intrusive like Clippy.

It is worth remembering that people have been creating animated characters, albeit hand drawn, for around one hundred years. Traditional cartoon characters seem to be readily accepted as being believable characters, the list of believable animated characters is almost endless. People can relate to characters like Bart Simpson, Mickey Mouse, Daffy Duck, etc. The quality of many of the characters being used by the research community to test anthropomorphic interface ideas are in the author's opinion primitive compared with the animated characters people see on television or in the cinema on a regular basis. This could be one of the reasons why sometimes people don't react well to characters used for anthropomorphic interface research. Given this, it is the author's personal opinion that closer ties with the arts community are required in order to create believable characters.

7.2.5 Acceptance by the research community

There is a great number of researchers advocating the benefits of anthropomorphic agents, some of these based on personal dreams of someday having conversational agents to help them with web searching and other everyday tasks (Hall). Others baked by research. However, opposition to these advocates is strong, Ben Shneiderman and Jaron Lanier are amongst the strongest critics of anthropomorphic agents.

Schneiderman (Sheiderman 1992) highlights that conversational anthropomorphic introduce unpredictability to the interface which is undesirable. I would contest that although it is true that we tend to associate anthropomorphism with unpredictability inherent in individualism it still doesn't rule out the usefulness of anthropomorphic interfaces. It is up to the research community to find ways of establishing trust on the agent from its user, I do not believe that the unpredictability inherent in anthropomorphic interfaces is unsurpassable. As a real life example we as humans depend on each other all the time, we form relationships where trust is built and we don't worry constantly that this trust might be breached.

Schneiderman (Sheiderman 1993) also worries about children interacting with anthropomorphic agents in educational software may lead to these children believing that they are automatons themselves. The author once again is sceptical, entirely from personal experience as a child he remembers playing with a pull cord Bugs Bunny that spoke and he doesn't ever remember believing the robotic Bugs was alive or that he worked on the same principle as it. From time to time there are stories in the newspapers of children believing they are Superman or the Power Rangers who end up hurting themselves or others, however, as we all know most children do not believe they are either Superman or the Power Rangers and these beliefs probably other problems and not a general problem with anthropomorphic representations. Similarly Laurel (Laurel 1997) also makes the point that the boundaries between humans and agents might become blurred. Laurel points out that if people get used to mistreating agents then they might begin to mistreat their real life agents (secretaries and personal assistants for example). This is something that the author has been concerned about for almost as long as he has been carrying out anthropomorphic agents research and believes it is an area worthy of investigation.

Lanier (Lanier 1995) goes as far to describe anthropomorphic agents as "wrong and evil". Evil because they force people to diminish themselves and wrong because they confuse the feedback that leads to good design. He states three main practical problems he envisages:

- If information consumers see the world through the eyes of agents then advertising will transform into the art of controlling agents, through bribing, hacking and other unsavoury methods. This is potentially a big problem, nowadays it is easy to ignore or block spam or pop-ups however if our personal agent filters all of our information we might have no way of knowing if someone is controlling our agent. However, once the law catches up with the technology, I suspect penalties for bribing or hacking agents will be severe.
- Since agents are computer programs they will have a lot more in common with each other than with humans. Agents would become the new information bottleneck, narrowing the richness of the Infobahn. He speculates that more likely than not an agent's model of what will interest you will be a cartoon model, and you will see a cartoon version of the world through the agent's eyes recreating the lowest common denominator model seen that plagues television, this would be a step back as the Infobahn was supposed to replace the broadcast model with something more inclusive.
- Finally Lanier says, "Agents will inevitably deliver an overdose of kitsch". As an example he says "Microsoft's Bob is the agent of the moment and it proposes to the user a life of caricature meaningless, sliding unintentionally into the grotesque". Dare I propose that in all probability many people have the exact same opinion of many television networks? I think this is a problem with media in general and the fact that some anthropomorphic agents "deliver an overdose of kitsch" is not a problem that is inherent in agents. I think this is something that has already been challenged by tasteful and successful interfaces like Ananova.

Lanier says that the major problem with anthropomorphic agents is that "Agents make people redefine themselves into lesser beings". He says that in order to make the agent seem smart users will change themselves, i.e. make themselves dumb. He sites the example of Apple Newton users, nowadays probably the same could be said of PocketPC and Palm users. He says that Newton users change the way they write and end up "contorting themselves" to write simple notes. I am of the opinion that although this is of some concern it is more to do with interface design and technology barriers than inherent problems with agents, nowadays PocketPCs can distinguish cursive text and it is no longer necessary to learn special symbols for different letters, although, users still need to adapt their writing in some way as special movements are still needed for line brakes and erasing characters for example. Lanier does point out that if agents were really autonomous and perhaps even conscious he would not have his objections would be disputable.

To summarize this section, let's say that the anti-agent community does propose issues that are of some concern and sometimes great concern. However, it is my opinion that these are not insurmountable and in fact make the area of seemingly intelligent anthropomorphic interfaces much more exiting.

7.3 Future work

In this section we discuss future work categorized into short term, medium term and long term.

7.3.1 Short Term

Further investigation and development of the rudeness rating method of evaluating interactive anthropomorphic characters:

We have seen how by analysing how much like a human users treat agents we can assess how much users like this agent. We will be developing an automated system for calculating the rudeness rating in order to facilitate experiments to expand and improve this method.

Attempt to identify characteristics of successful agents in general use today:

Now that there are many anthropomorphic interfaces in general use it would be a good idea to build a record of these in order to identify which aspects or characteristic the successful and unsuccessful characters share, if any.

7.3.2 Medium Term

Begin to explore how users react to agents that seem to react emotionally to user input.

We have seen how users clearly prefer agents which exhibit accurate emotional reactions, however, we haven't explored how users react to agents which react emotionally to their input, for instance how would a user feel about an agent that seems to grow fond of the user or an agent which is just the contrary. We have already gone some way towards designing a setting to carry out some of these experiments.

Ideally these experiments would set up online as it makes it easier to obtain large numbers of users. We plan to set up some simple online interactions with two groups of users; A control group interact with an agent with a fixed neutral expression versus an experimental group interacting with agents that display more expressive, human-like behaviour. The groups will be decided by looking at IP addresses and using a mathematical formula to assign them to the control or experimental group. The interaction with the agent will be similar to the interaction in the pilot experiment. This extends upon the interface used for the two-way interaction experiments, i.e. a web searching assistant.

The agent should be capable of displaying emotions reacting to the interaction with the user and the results of their queries. In interacting with the user the agent executes a three step loop: sense (listen to the user), think (decipher the user input), act (react to it accordingly). The user interacts with the agent using natural language. This is done via keyboard input, within a web page using a familiar search engine interface such as Google. To decipher the natural language input we are using a natural language engine which is a servlet running in our web server. The reactions of the agent could be of a few different types:

- Produce a list of links.
- Produce one link.
- Ask for confirmation.
- Reply query.

We would have two agents, one that is completely neutral with a fixed facial expression, and the other that with a 'reduced' emotion set. This expression set could comprise of:

- **Happy:** The agent might be happy when it successfully carries out tasks for the user. Or whenever it is praised by it.

- **Sad:** The agent might feel sad when reprimanded by the user or when it fails to find useful information for the user.
- **Fear:** The agent could fear being reprimanded by the user.
- **Hope:** The agent could hope to be successful at a task. Or it could hope to be praised.
- **Proud:** The agent might feel proud when it has been praised. Or when it has carried out successful queries.
- **Shame/Embarrassment:** The agent might feel ashamed when it has failed to understand the user or it has failed to produce adequate answers.
- **Anger:** The agent might feel angry when the user unfairly reprimands it.
- **Love:** The agent might grow to love the user maybe?
- **Hate:** The agent might grow to hate the user if it is constantly abused.
- **Amused:** The agent might be amused by the user when or it might amuse itself by saying something funny.

The simplified model of emotion for the action/event reaction of these emotions is as follows:

1) Type of Event: User makes an intelligible query.

Action: Agent makes a query, which seems to return a perfect match.
Emotion: Pride, Happy.

Action: Agent makes a query, which returns a good set of matches.
Emotion: Pride, Happy.

Action: Agent makes a query, which returns a bad set of links.
Emotion: Shame, Sad.

Action: Agent makes a query, which doesn't return any matches.
Emotion: Shame, Sad.

2) Type of Event: User makes an unintelligible query.

Action: Agent asks user to rephrase query.

Emotion: Shame/Embarrassment.

3) Type of Event: User praises agent.

Action: Agent thanks users and lets him know that it is glad it has been praised.
Emotion: Pride.

Action: Agent thanks user in a funny way.
Emotion: Amused, Happy, Proud.

4) Type of Event: User reprimands agent after a bad set of matches or no matches.

Action: Agent apologises to the user.
Emotion: Shame, Sad, Hope that next query will be more successful, Fear of the next query producing a bad set of matches or no matches at all.

5) Type of Event: User reprimands agent after a good set of matches or almost a perfect match.

Action: Agent apologises to the user.
Emotion: Shame, Sad, Hope that next query will be more successful.

6) Type of Event: User reprimands agent after a good set of matches or almost a perfect match for a second time in the last 5 cycles.

Action: Agent apologises, not sounding too sincere.
Emotion: Angry, Sad.

6) Type of Event: User praises agent 5 times in the previous 5 cycles.

Action: Agent says funny things or flirts with the user and lets them know how pleased he/she is.
Emotion: Love.

Further development of the rudeness rating and the development of an automated system for calculating this will help to analyse data for these experiments.

7.3.3 Long Term

To study user's reactions to intelligent and conscious machines; VICKI – Vociferous Intelligent Computer Knowledge Interface:

Given the science fiction view that people will show great resistance to intelligent machines possibly because their own humanity is challenged it would be an interesting exercise to study how people would react to a computer than seems not only intelligent but also aware of it's consciousness.

This is a study that will require great care and preparation, the design, setting and implementation need to be carefully considered, as users need to be successfully tricked into thinking that they are interacting with a an intelligent and conscious machine.

The setting of the experiment needs to match the expectations of test subjects, for example, test subject will not believe that they are interacting with an intelligent conscious computer if they are sat in front of an old PC in an open computer lab. The computer needs to appear to be intelligent and conscious of its existence; therefore the experiment would certainly be a Wizard of Oz type experiment, i.e. the computer is controlled by a human.

The intelligent computer itself needs to be believable; I propose that not having an anthropomorphic representation of the machine at all is the best option. We have seen in this research how users have different expectations of different looking anthropomorphic representations therefore to get rid of this variable perhaps the beast way is to not to have this representation at all. Something akin to the look of HAL in 2001 Space Odyssey is possibly a good idea, a high-tech looking console with a computer eye (camera) which the users believe is for the machine to see them but in reality it is for the wizard to see them.

There are many challenges is designing and carrying out this experiment. At the end of the experiment the truth will be revealed to users, however considering the nature of this experiment many users might experience many strong emotions and they might feel used and cheated, this needs to be taken into consideration. There are practical and issues as well; obtaining a setting that looks like a realistic secure facility could prove a challenge. The experimenters could perhaps be trained actors as real researchers might not be believable in their fake roles. Obtaining users is also an issue that needs to be considered carefully, users need to believe that they are being presented with a once in a lifetime opportunity to interact with the best most intelligent computer in the world, the first successful attempt at artificial life. If users just answer a dull poster and then are presented with this great opportunity they might suspect that something isn't quite what it seems, if on the other hand they feel they were hand picked somehow then they might be more willing to believe the situation is real. Users should not be computer scientists or even computer literate, people who keep up with computer science research know that we are many years from a truly intelligent computer, in fact we might never reach that goal.

Finally, what are we hoping to obtain from this experiment? Several things, one is to obtain data on how people will behave towards what they believe is the first artificial life ever created. Secondly how this makes them feel, do they feel as though their humanity is somehow diminished given that machines are conscious now? Are they happy to make friends with the machine?

References

Ananova-Birthday-Report. *Ananova celebrates her first birthday.* http://www.ananova.com/about/story/sm_263306.html.

Ananova-URL. *Ananova.* www.ananova.com.

André, E. (2000). Enhancing Adaptive Hypermedia Presentation Systems by Lifelike Synthetic Characters. AH2000, LNCS 1892. pp. 1-4

André, E., M. Klesen, et al. (1999). Integrating models of personality and emotions into lifelike characters. In Proceedings International Workshop on Affect in Interactions. Towards a New Generation of Interfaces. pp.

André, E. and T. Rist (2000). Presenting Through Performing: On the Use of Multiple Lifelike Characters in Knowledge-Based Presentation Systems. Second International Conference on Intelligent User Interfaces (IUI 2000). pp. 1-8

André, E., T. Rist, et al. (2000). *The Automated Design of Believable Dialogues for Animated Presentation Teams.* Embodied Conversational Agents.

Ball, G., D. Ling, et al., Eds. (1997). Lifelike Computer Characters: the Persona project at Microsoft Research. Software agents. London, AAAI Press.

Bates, J. (1994). *The Role of Emotion in Believable Agents.* Communications of the {ACM} **37**(7): 122--125.

Begntsson, B., J. K. Burgoon, et al. (1999). The Impact of Anthropomorphic Interfaces on Influence, Understanding, and Credibility. Thirty-Second Annual Hawaii International Conference on System Sciences, Maui, Hawaii, IEEE. pp.

Brin, S. and L. Page (1998). *The anatomy of a large-scale hypertextual Web search engine.* Computer Networks and ISDN Systems **30**(1-7): 107-117.

Buck, J. (1999). Interactive Communicative Humanoid Agent for Movie Selection. Behavior Planning for Life-Like Characters and Avatars
In Conjunction with the i3 Spring Days '99, Sitges, Spain. pp.

Cassell, J., Ed. (2002). Embodied Conversation: Integrating Face and Gesture into Automatic Spoken Dialogue Systems. Spoken Dialogue Systems.

Cassell, J., T. Bickmore, et al. (1999). Conversation as a System Framework: designing Embodied Conversational Agents. Embodied Conversational Agents. Cambridge, MA, MIT Press.

Cassell, J., T. Bickmore, et al. (2001). *More Than Just a Pretty Face: Conversational Protocols and the Affordances of Embodiment.* Knowledge-Based Systems(14): 55-64.

Cassell, J., C. Pelachaud, et al. (1994). *ANIMATED CONVERSATION Rule-based Generation of Facial Expression.* Computer Graphics **28**: 413--420.

Cassell, J. and K. R. Thórisson (1999). *The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents.* Applied Artificial Intelligence(13): 519-538.

Cramer, D. (1998). Fundamental Statistics for Social Research, Routledge.

Damper, R., W. Hall, et al., Eds. (1994). Multimedia Technologies and Future Applications. London, Pentech Press.

Dick, P. (1968). Do Androids Dream of Electric Sheep?, Random House, Inc.

Dohi, H. and M. Ishizuka (1997). Visual Software Agent: A Realistic Face-to-Face Style Interface connected with WWW/Netscape. Proceedings of IJCAI-97 Workshop on Intelligent Mutlimodal Systems, Nagoya. pp. 17-22

Dohi, H. and M. Ishizuka (1999). Visual Software Agent Interface with a User-tracking Flat Panel Display for Continuous Eye Contact. 3rd World Multiconference on Systems, Cybernetic and Information and 5th Int'l Conf. on Information Systems Analysis and Synthesis (SCI/ISAS'99), Orlando, Florida, USA. pp.

Don, A., S. Brennan, et al. (1992). Anthropomorphism: From Eliza to Terminator. Proceedings of CHI-92, Monterey, CA. pp. 67-70

Efron, D. (1972). Gesture, Race and Culture, The Hague: Mouton & Co.

Ekman, P. (1993). *Facial Expression and Emotion.* American Psychologist **48**(4): 384-392.

Ekman, P. (1999). Basic Emotions. Handbook of Cognition and Emotion. M. Power. Sussex, UK, John Wiley & Sons, Ltd.

Ekman, P. and W. Friesen (1975). Unmasking the Face: a guide to recognizing emotions from facial cues, Prentice-Hall, Inc.

Elliott, C. (1994). Multi-Media communication with emotion-driven 'believable agents'. AAAI Technical Report for the Spring Symposium on Believable Agents, AAAI, American Association for Artificial Intelligence. pp. 16-20

Elliott, C. and E. Melchoir (1995). *Getting to the point: Emotion as a necessary and sufficient element of story construction.*Elliott, C. and E. Melchoir,Stanford University,pp.

Foner, L. (1993). *What's An Agent, Anyway? A Sociological Case Study - Technical Report.*Foner, L.,

Hall, W. *Professor Wendy Hall's Inaugural Lecture.* http://godzilla.ecs.soton.ac.uk/~km/streams/wendy-inaugural.xdm.

Haptek-URL. *Haptek.* www.haptek.com.

Hopper, D. I. (2000). Virtual news jock Ananova goes live. CNN.com.

Kamm, C., M. Walker, et al. (1997). The role of speech processing in human-computer intelligent communication. NSF Workshop on human-centered systems: Information, interactivity, and intelligence. pp.

Kendon, A. (1999). An Agenda for Gesture Studies. Semiotic Review of Books. **Volume 7 (3)**.

Koda, T. (1996). Agents with Faces: A Study on the Effects of Personification of Software Agents. MIT Media Laboratory. Boston, MIT.

Koda, T. and P. Maes (1999). Agents with Faces: The Effects of Personification of Agents. Proceedings of 3rd World Multiconference on Systems, Cybernetic and Information and 5th Int'l Conf. on Information Systems Analysis and Synthesis (SCI/ISAS'99), Orlando, Florida, USA. pp. 242-247

Lachman, R. (1997). Animistic Interface: Experiments in Mapping Character Animation to Computer Interface, MIT Press.

Lanier, J. (1995). *Agents of Alienation.* Journal of Consciousness Studies(2): 76-81.

Lanier, J. (1996). My Problem with Agents. Wired Magazine**:** 43.

Lau, R., G. Flammia, et al. (1997). WebGalaxy: Beyond Point and Click - A Conversational Interface to a Browser. Sixth International World Wide Web Conference, Santa Clara. pp. 119-128

Laurel, B. (1997). Interface Agents: Metaphors with Character, AAAI Press / The MIT Press.

Lee, Y., D. Terzopoulos, et al. (1995). Realistic Modelling for Facial Animation. 22nd annual ACM conference on Computer graphics, Los Angeles, CA USA. pp.

Maes, P. (1996). Artificial Life Meets Entertainment: Lifelike Autonomous Agents. Clicking In, Hot Links to a Digital Culture. L. H. Leeson. Seattle, Bay Press.

Maldonado, H., A. Picard, et al. (1998). Tigrito: a multi-mode interactive improvisational agent. International Conference on Intelligent User Interfaces, San Francisco, California, United States, ACM Press. pp. 29-32

Mateas, M. (1997). *An Oz-Centric Review of Interactive Drama and Belivable Agents.*Mateas, M.,School of Computer Science, Carnegie Mellon University,pp.

McCloud, S. (1994). Understanding Comics The Invisible Art, HarperPerennial.

Microsoft (1999). Microsoft Agent Software Development Kit, Microsoft Press.

Microsoft-Press-Release. *Farewell Clippy: What's Happening to the Infamous Office Assistant in Office XP.* http://www.microsoft.com/presspass/features/2001/apr01/04-11clippy.asp.

Mulder, A. (1996). *Hand Centered Studies of Human Movement Project.*Mulder, A.,School of Kinesiology, Simon Fraser University,pp.

Mylopoulos, J., L. Chung, et al. (2001). *Exploring Alternatives During Requirements Analysis.* IEEE **18**(1): 92-96.

Nass, C. and K. M. Lee (2000). Does computer-generated speech manifest personality? an experimental test of similarity-attraction. Conference on Human Factors and Computing Systems, The Hague, The Netherlands, ACM Press. pp.

Nass, C., J. Steuer, et al. (1994). Computers are social actors. Conference proceedings on Human factors in computing systems, Boston, Massachusetts, United States, ACM Press. pp. 72-78

Olive, J. (1997). "The Talking Computer": Text to Speech Synthesis. Hal's Legacy: 2001's Computer as Dream and Reality. D. Stork. Cambridge, MA, MIT Press.

Ortony, A., G. Clore, et al. (1988). The Cognitive Structure of Emotions, Cambridge University Press.

Pelachaud, C., N. I. Badler, et al. (1996). *Generating Facial Expressions for Speech.* Cognitive Science **20**(1): 1-46.

Persona-Project. *Persona Project.* http://research.microsoft.com/ui/persona/home.htm.

Power, G., R. Damper, et al. (2002). Realism and Naturalness in a Conversational Multimodal Interface. ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, Kloster Irsee, Germany. pp.

Power, G., G. Will, et al. (2002). User Perception of Anthropomorphic Characters with Varying Levels of Interaction. The 16th British HCI Group Annual Conference, London, United Kingdom. pp.

Ralston, J. V., D. B. Pisoni, et al. (1995). Perception and comprehension of speech. Applied Speech Technology, Boca Raton: CRC Press. pp. 233-288.

Reeves, B. and C. Nass (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places, Cambridge University Press.

Rimé, S. (1991). Gesture and Speech. Fundamentals of Nonverbal behaviour. Gesture and Speech, Cambridge University Press. pp. 239-281.

Ross, F. D. (1981). *The aprosodias: functional - anatomical organization of affective components of langauge in the right hemisphere.* Archives of Neurology(38): 561-569.

Ruttkay, Z. and H. Noot (2000). Animated CharToon faces. first international symposium on Non-photorealistic animation and rendering, Annecy, France. pp. 91-100

Ruttkay, Z., H. Noot, et al. (2000). CharToon Faces for the Web. WWW9 Poster Proceedings. pp.

Sheiderman, B. (1992). Designing the User Interface: Strategies for Effective Human-Computer Interaction, Addison Wesley.

Sheiderman, B. (1993). A non Anthropomorphic style guide: overcoming the Humpty Dumpty Syndrome, Ablex Publishing.

Sheth, B. (1994). A Learning Approach to Personal Information Filtering. Boston, Massachusetts, USA, Massachusetts Institute of Technology.

Siegel, S. and N. J. Castellan (1988). Nonparametric Statistics for the behavioral scinences, McGRAW-HILL.

Takama, H. Dohi, et al. (1998). A Visual Anthropomorphic Agent with Learning Capability of Cooperative Answering Strategy through Speech Dialog. Asia Pacific Computer Human Interaction (APCHI'98), Japan. pp. 260-265

Thomas, F. and O. Johnson (1981). Disney Animation: The Illusion of Life. New York, Abbeville Press.

Thórisson, K. R. and J. Cassell (1996). Why Put An Agent In a Body: The Importance of Communicative Feedback in Human-Humanoid Dialogue. Lifelike Computer Characters, Utah, USA. pp.

Trott, B. (1998). Microsoft's paper-clip assistant killed in Denver. CNN.com.

Trower, T. (1997). Creating Conversational Interfaces for Interactive Software Agents. CHI 97, Electronic Publications: Tutorials. pp.

Velásquez, J. (1997). Modeling Emotions and Other Motivations in Synthetic Agents. Proceedings of AAAI-97. pp.

Verbot-URL. *Virtual Personalities Inc.* www.verbot.com.

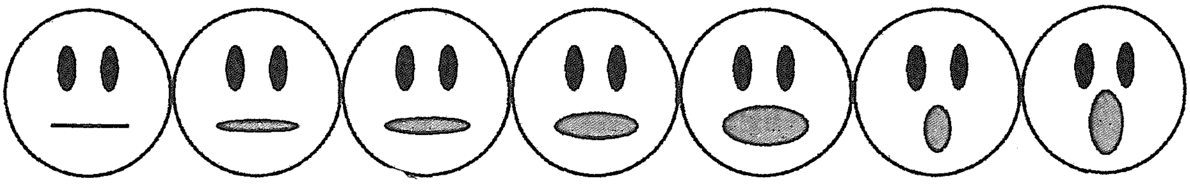
Walker, J. H., L. Sproull, et al. (1994). Using a Human Face in an Interface. Human Factors in Computing Systems CHI'94, New York, ACM Press. pp. 85-91

Zue, V. (1999). Talking with Your Computer. Scientific American: 40-41.

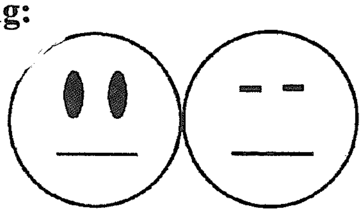
Appendix 1

Smiley face character A:

Frames used for speaking:



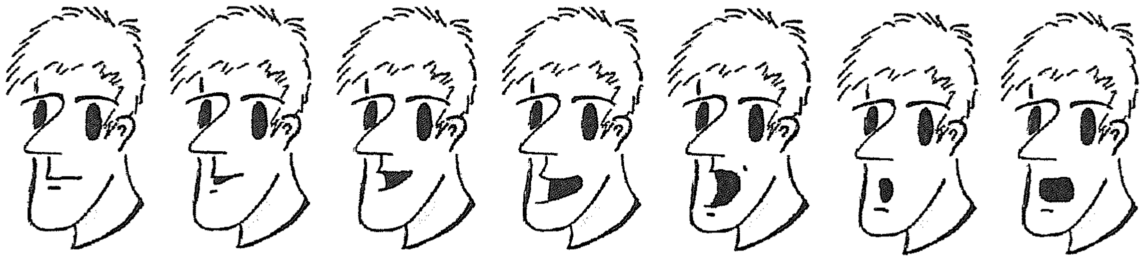
Frames used for speaking:



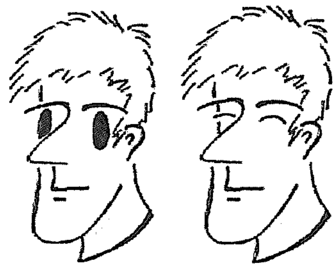
Frame used for smiling:

Cartoon character B:

Frames used for speaking:



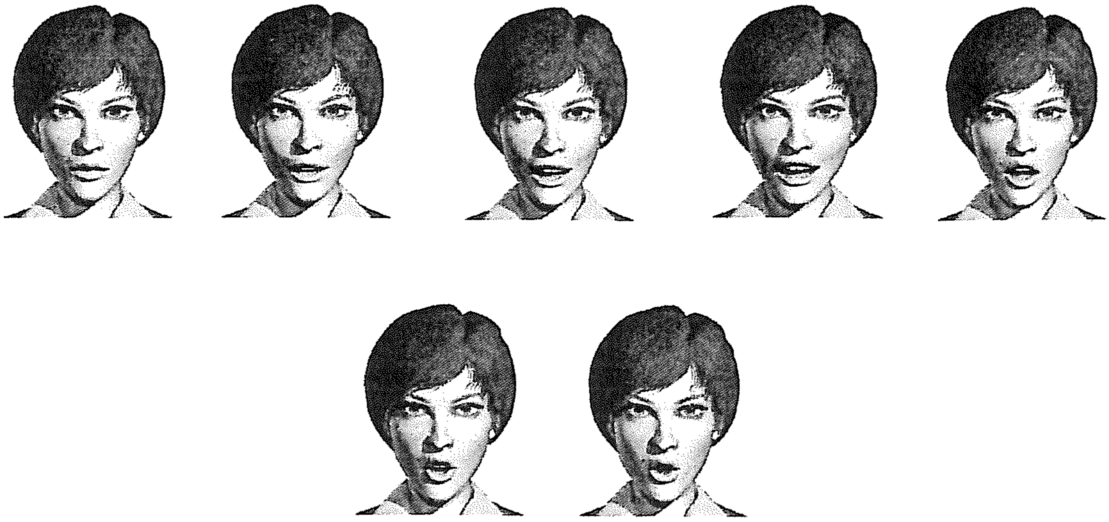
Frames used for blinking:



Frame used for smiling:

Realistic character C:

Frames used for speaking:

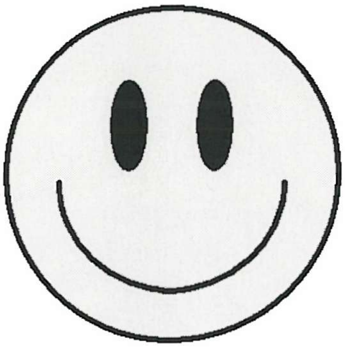


Frames used for blinking:

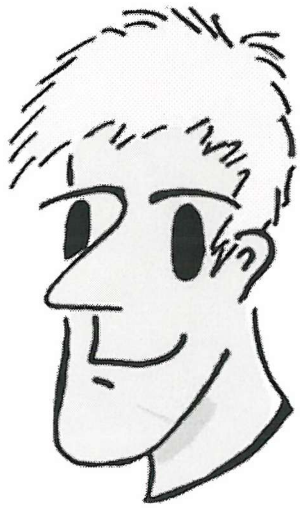
Frame used for smiling:

Appendix 2

Rate these faces in a scale of 1-7 for the following categories (1 for a low rating and 7 for a high rating):



A



B

C

	Unfriendly						**Friendly**
	1	2	3	4	5	6	7
A							
B							
C							

	Unintelligent						**Intelligent**
	1	2	3	4	5	6	7
A							
B							
C							

	Unpleasant						**Pleasant**
	1	2	3	4	5	6	7
A							
B							
C							

	Uninteresting						**Interesting**
	1	2	3	4	5	6	7
A							
B							
C							

	Unhelpful						Helpful
	1	2	3	4	5	6	7
A							
B							
C							

Performance questions (1 Strongly disagree, 7 Strongly agree)

Produced accurate matches.

	Strongly Disagree						Strongly Agree
	1	2	3	4	5	6	7
A							
B							
C							

Understood what I wanted.

	Strongly Disagree						Strongly Agree
	1	2	3	4	5	6	7
A							
B							
C							

Appendix 3

Instructions for test subjects:

- This is a helper character that has been designed to help you search for WebPages, give you lists of links and take you to websites.
- Just talk to it naturally and ask for what you want. For example if you want to visit Yahoo! Just say something like: "Load Yahoo!"
- Before starting a new search close all the browser windows and the software will get confused.

Appendix 4

I A M

Facial survey

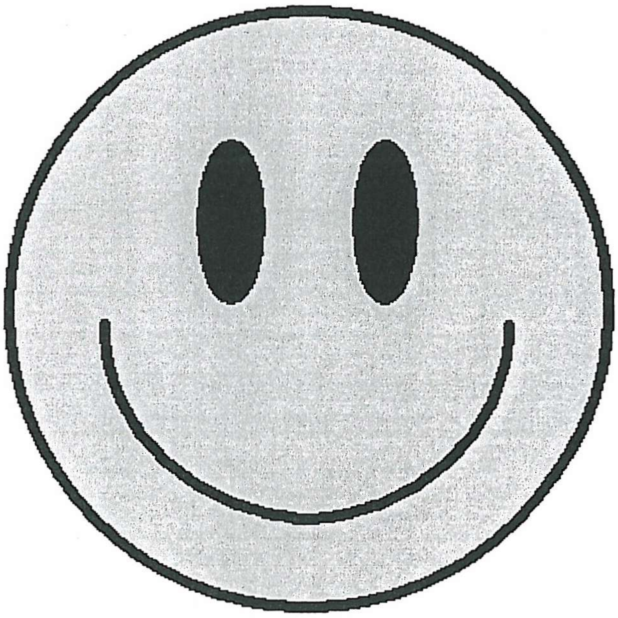
E-Mail Address

Gender

Male Female

ge <18 18-25 26-30 31-35 36-40 41-45 46-50 51-65 >55

Rate these three faces in a scale of 1 to 7 for the following categories:



Unfriendly 1 2 3 4 5 6 7 Friendly

Unintelligent 1 2 3 4 5 6 7 Intelligent

Unpleasant 1 2 3 4 5 6 7 Pleasant

Uninteresting 1 2 3 4 5 6 7 Interesting

Unfriendly ◌ 1 ◌ 2 ◌ 3 ◌ 4 ◌ 5 ◌ 6 ◌ 7 Friendly

Unintelligent ◌ 1 ◌ 2 ◌ 3 ◌ 4 ◌ 5 ◌ 6 ◌ 7 Intelligent

Unpleasant ◌ 1 ◌ 2 ◌ 3 ◌ 4 ◌ 5 ◌ 6 ◌ 7 Pleasant

Uninteresting ◌ 1 ◌ 2 ◌ 3 ◌ 4 ◌ 5 ◌ 6 ◌ 7 Interesting

Unfriendly ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 Friendly

Unintelligent ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 Intelligent

Unpleasant ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 Pleasant

Uninteresting ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 Interesting

Comments [ ] SUBMIT

Guillermo Power
IAM Research Group
Department of Electronics and Computer Science
University of Southampton
Highfield, Southampton SO17 1BJ, United Kingdom
gp98r@ecs.soton.ac.uk

Appendix 5

No-interaction

$$\beta = Z\left(\frac{k(k+1)}{6N}\right)^{1/2}$$

p	Z	β
0.05	2.394	0.618
0.10	2.128	0.549
0.15	1.96	0.506
0.20	1.834	0.4735
0.25	1.732	0.447

Friendliness:

$|\overline{R}_A - \overline{R}_B| = 1.16$ >95%

$|\overline{R}_A - \overline{R}_C| = 1.63$ >95%

$|\overline{R}_B - \overline{R}_C| = 0.471$ >75%

Intelligence:

$|\overline{R}_A - \overline{R}_B| = 0.50$ >80%

$|\overline{R}_A - \overline{R}_C| = 1.15$ >95%

$|\overline{R}_B - \overline{R}_C| = 0.65$ >95%

Pleasantness:

$|\overline{R}_A - \overline{R}_B| = 0.98$ >95%

$|\overline{R}_A - \overline{R}_C| = 1.06$ >95%

$|\overline{R}_B - \overline{R}_C| = 0.08$ (no difference)

Interestingness:

$|\overline{R}_A - \overline{R}_B| = 0.59$ >90%

$|\overline{R}_A - \overline{R}_C| = 0.52$ >85%

$|\overline{R}_B - \overline{R}_C| = 0.07$ (no difference)

One-way-interaction

$$\beta = Z\left(\frac{k(k+1)}{6N}\right)^{1/2}$$

p	Z	β
0.05	2.394	0.3809
0.10	2.128	0.3386
0.15	1.96	0.31186

Friendliness:

$|\overline{R}_A - \overline{R}_B| = 0.45$ >95%

$|\overline{R}_A - \overline{R}_C| = 0.84$ >95%

$|\overline{R}_B - \overline{R}_C| = 0.39$ >95%

Intelligence:

$|\overline{R}_A - \overline{R}_B| = 0.61$ >95%

$|\overline{R}_A - \overline{R}_C| = 0.77$ >95%

$|\overline{R}_B - \overline{R}_C| = 0.16$ (no difference)

Pleasantness:

$|\overline{R}_A - \overline{R}_B| = 0.03$ (no difference)

$|\overline{R}_A - \overline{R}_C| = 0.60$ >95%

$|\overline{R}_B - \overline{R}_C| = 0.63$ >95%

Interestingness:

$|\overline{R}_A - \overline{R}_B| = 0.40$ >95%

$|\overline{R}_A - \overline{R}_C| = 0.32$ >85%

$|\overline{R}_B - \overline{R}_C| = 0.08$ (no difference)

Helpfulness:

$|\overline{R}_A - \overline{R}_B| = 0.20$ (no difference)

$|\overline{R}_A - \overline{R}_C| = 0.16$ (no difference)

$|\overline{R}_B - \overline{R}_C| = 0.04$ (no difference)

Two-way-interaction

$$\beta = Z\left(\frac{k(k+1)}{6N}\right)^{1/2}$$

p	Z	β
0.05	2.394	1.0208
0.10	2.128	0.9073
0.15	1.96	0.8357
0.20	1.834	0.7820
0.25	1.732	0.7385

Friendliness:

$\left|\overline{R}_A - \overline{R}_B\right| = 0.14$ (no difference)

$\left|\overline{R}_A - \overline{R}_C\right| = 1.04$ >95%

$\left|\overline{R}_B - \overline{R}_C\right| = 0.90$ >90%

Intelligence:

$\left|\overline{R}_A - \overline{R}_B\right| = 0.50$ (no difference)

$\left|\overline{R}_A - \overline{R}_C\right| = 1.00$ >90%

$\left|\overline{R}_B - \overline{R}_C\right| = 0.50$ (no difference)

Pleasantness:

$\left|\overline{R}_A - \overline{R}_B\right| = 0.23$ (no difference)

$\left|\overline{R}_A - \overline{R}_C\right| = 0.73$ >75%

$\left|\overline{R}_B - \overline{R}_C\right| = 0.50$ (no difference)

Interestingness:

$\left|\overline{R}_A - \overline{R}_B\right| = 0.41$ (no difference)

$\left|\overline{R}_A - \overline{R}_C\right| = 0.55$ (no difference)

$\left|\overline{R}_B - \overline{R}_C\right| = 0.14$ (no difference)

Helpfulness:

$\left|\overline{R}_A - \overline{R}_B\right| = 0.19$ (no difference)

$\left|\overline{R}_A - \overline{R}_C\right| = 0.04$ (no difference)

$\left|\overline{R}_B - \overline{R}_C\right| = 0.19$ (no difference)

Accuracy:

$\left|\overline{R}_A - \overline{R}_B\right| = 0.22$ (no difference)

$\left|\overline{R}_A - \overline{R}_C\right| = 0.59$ (no difference)

$\left|\overline{R}_B - \overline{R}_C\right| = 0.37$ (no difference)

Understanding:

$\left|\overline{R}_A - \overline{R}_B\right| = 0.64$ >0.70%

$\left|\overline{R}_A - \overline{R}_C\right| = 0.59$ (no difference)

$\left|\overline{R}_B - \overline{R}_C\right| = 0.05$ (no difference)

Voices

$$\beta = Z\left(\frac{k(k+1)}{6N}\right)^{1/2}$$

p	Z	β
0.05	2.394	0.3854
0.10	2.128	0.3426
0.15	1.96	0.3156
0.20	1.834	0.2953
0.25	1.732	0.2789

Friendliness:

$|\overline{R}_A - \overline{R}_B| = 0.48$ >95%

$|\overline{R}_A - \overline{R}_C| = 0.06$ (no difference)

$|\overline{R}_B - \overline{R}_C| = 0.42$ >95%

Intelligence:

$|\overline{R}_A - \overline{R}_B| = 0.27$ >70%

$|\overline{R}_A - \overline{R}_C| = 0$ (no difference)

$|\overline{R}_B - \overline{R}_C| = 0.27$ >70%

Pleasantness:

$|\overline{R}_A - \overline{R}_B| = 0.31$ >85%

$|\overline{R}_A - \overline{R}_C| = 0.22$ >70%

$|\overline{R}_B - \overline{R}_C| = 0.09$ (no difference)

Helpfulness:

$|\overline{R}_A - \overline{R}_B| = 0$ (no difference)

$|\overline{R}_A - \overline{R}_C| = 0.04$ (no difference)

$|\overline{R}_B - \overline{R}_C| = 0.04$ (no difference)

Understanding:

$|\overline{R}_A - \overline{R}_B| = 0.62$ >0.95%

$|\overline{R}_A - \overline{R}_C| = 0.02$ (no difference)

$|\overline{R}_B - \overline{R}_C| = 0.64$ >95%

Appendix 6

		Smiley		Man		Woman
User 2	Keyword	wall climbing. (wall+climbing)false	Keywords	robin williams. (robin+williams)false	Keywords	brookside. (brookside)false
	Keyword	jumping. (jumping)false	*	robin williams singer. (robin+williams+singer)false	Keywords	friends. (fnends)false
	Question	does ncr have a branch in korea. (does+ncr+have+a+branch+in+korea)false	Keywords	westlife. (westlife)false	Keywords	american drama friends. (american+drama+friends)false
	*	does ncr have a branch in korea?. (does+ncr+have+a+branch+in+korea)false	*	westlife. (westlife)false	Keywords	walking. (walking)false
	*	does ncr have a branch in korea?. (does+ncr+have+a+branch+in+korea)false	Keywords	westlife five boys from irish. (westlife+five+boys+from+irish)false		
	Keyword	korea. (korea)false	Complex	find a paper about scatter/gather method in text classification. links(a+paper+about+scatter/gather+method+in+text+classification)true		
	*	does ncr have a branch in korea?. (does+ncr+have+a+branch+in+korea)false	Keywords	threading electronic message. (threading+electronic+message)false		
User 3	Simple	i want blind guardian videos. (i+want+blind+guardian+videos)false	Question	where is a good XSLT reference?. (where+is+a+good+xslt+reference)false	Keywords	snowboarding in aviemore. (snowboarding+in+aviemore)false
	Simple	i want cds of the gathering. (i+want+cds+of+the+gathering)false	Complex	a book on artificial intelligence using c++. (a+book+on+artificial+intelligence+using+c++)false	Question	where can i flyfish in scotland. (where+can+i+flyfish+in+scotland)false
	Keywords	music by the gathering. (music+by+the+gathering)false	Question	when will the lored of the rings be released. (when+the+lored+of+the+rings+be+released)false	Question	where is the offices of flyshop.co.za. (where+is+the+offices+of+flyshop+co+za)false
	Keywords	the jerry springer show. *(the+jerry+springer+show)*false	*	when will the lord of the rings be released. (when+the+lord+of+the+rings+be+released)false	*	where are the offices of flyshop.co.za. (where+are+the+offices+of+flyshop+co+za)false
	Keywords	the big breakfast show. (the+big+breakfast+show)false				
	Keywords	new season of that 70s show. (new+season+of+that+70s+show)false				
User 4	Keywords	britney spears. (britney+spears)false	Keywords	star wars figurines. (star+wars+figurines)false	Keywords	mobile agents. (mobile+agents)false
	Question	why is britney spears so popular ?. (why+is+britney+spears+so+popular)false	Simple	find me star war hobbies. links(star+war+hobbies)true	Keywords	mobile agent security. (mobile+agent+security)false
	Keywords	cranberry lyrics. (cranberry+lyrics)false	Simple	star wars role playing games. (star+wars+role+playing+games)false	Keywords	trust propagation. (trust+propagation)false
	Keywords	*cranberries lyrics.* (cranberries+lyrics)false	NL	*making it work is not useful.* (making+it+work+is+not+useful)false	Keywords	big brother tv show. (big+brother+tv+show)false
	Keywords	cranberries lyrics. (cranberries+lyrics)false	Keywords	palm pilot auction. (palm+pilot+auction)false		
	Keywords	craberries group members. (craberries+group+members)false	Keywords	cheap palm pilots. (cheap+palm+pilots)false		
	Keywords	cranberries group members. (cranberries+group+members)false	*	cheap palm pilots. (cheap+palm+pilots)false		
	Keywords	best song written by cranberries. (best+song+written+by+cranberries)false	Keywords	r2d2. (r2d2)false		
	Keywords	cranberries top songs. (cranberries+top+songs)false				
	Keywords	cranberries . (cranberries)false				
	Keywords	badminton. (badminton)false				
	Keywords	badminton ardy wiranata. (badminton+ardy+wiranata)false				
	Keywords	tennis. (tennis)false				
	Keywords	badminton world rankings. (badminton+world+rankings)false				
	Keywords	badminton thomas cup. (badminton+thomas+cup)false				
	Keywords	badminton razif sidek. (badminton+razif+sidek)false				
	*	badminton razif sidek. (badminton+razif+sidek)false				
User 5			Question	where can i buy frasier memorabilia online. (where+can+i+buy+frasier+memorabilia+online)false	Keywords	regine. (regine)false
			Question	which company sells the cheapest portable iron?. (which+company+sells+the+cheapest+portable+iron)false	Keywords	regine pictures. pictures(regine)true
					Keywords	regine velasquez. (regine+velasquez)false
					Keywords	regine. (regine)false
User 6	Chichat	hello. (hello)false	Complex	Please find some sites to do with Amorphous Computing. I suggest looking in links(find+sites+to+do+with+amorphous+computing+i+suggest+looking+in+the+area+of+ubiquit	Complex	Tell me a little about squash.. (tell+me+a+little+about+squash)false
	Question	Could you find me some sites about the music band c links(the+music+band+called+fields+of+the+nephilim)true	Question	What is complexity theory?. (what+is+complexity+theory)false	Question	Could you find me some links on the game of squash please?. links(the+game+of+squash)true
	NL	Those were interesting, but could you find any sites cc links(those+were+interesting+but+find+any+sites+containing+similar+bar	Question	Is anyone working with large numbers of nodes that are spread over a common (is+anyone+working+with+large+numbers+of+nodes+that+are+spread+over+a+common+mediu	Question	What are the upcoming events in the game of Squash?. (what+are+the+upcoming+events+in+the+game+of+squash)false
	Question	What is the band Cranes currently doing?. (what+is+the+band+cranes+currently+doing)false	Simple	Find stuff on ancient mythology.. links(ancient+mythology)true	Question	Can you find me some pictures of Penfold?. pictures(can+find+me+pictures+of+penfold)false
	Chitchat	Why are you smiling?. (why+are+smiling)false			Question	What cartoon did Penfold star in?. (what+cartoon+did+penfold+star+in)false
	Question	Do you know of a device that bakes bread?. (do+know+of+a+device+that+bakes+bread)false			Simple	Find some sites about DangerMouse. links(dangermouse)true
	Question	Is there a product that automatically bakes bread?. (is+there+a+product+that+automatically+bakes+bread)false				
	Question	Can you find some place that sells a "Williams-Sonoma Bread Machine" for a cheap price?. (can+find+place+that+sells+a+"williams-sonoma bread machine"+for+a+cheap+price)false				
	Keywords	bread cooking machine.				

		(bread+cooking+machine)false				
	Question	Is "Tool" still going or has it all stopped now that "A Perfect Circle" is formed?. (is+"tool"+still+going+or+has+it+all+stopped+now+that+"a perfect circle"+is+formed)false				
	Question	Does the band "A Perfect Circle" mean the end of "Tool"?. (does+the+band+"a perfect circle"+mean+the+end+of+"tool")false				
	Complex	Find me the relationship between Tool and A Perfect Circle.. links(the+relationship+between+tool+and+a+perfect+circle)true				
	Question	Who is Maynard?. (who+is+maynard)false				
User 7	Keywords	mercedes remote control cars. (mercedes+remote+control+cars)false			Keywords	dream team. (dream+team)false
	Keywords	ipswich town football club statistics. (ipswich+town+football+club+statistics)false			Keywords	dream team sky. (dream+team+sky)false
	Keywords	ipswich town football club statistics. (ipswich+town+football+club+statistics)false			Keywords	scholary representations. (scholary+representations)false
					*	scholarly representations. (scholarly+representations)false
					Keywords	scholarly modeling. (scholarly+modeling)false
User 8	Simple	show me some hypertext authoring links. links(show+me+hypertext+authoring)true	Simple	show me a mountain biking webiste. (show+me+a+mountain+biking+webiste)false	Simple	I'd like to see a website about guitars. website(i+d+like+to+see+a+website+about+guitars)false
	Simple	i need to know about the hypertext authoring process. * (i+need+to+know+about+the+hypertext+authoring+process)false		show me a mountain biking website. website(show+me+a+mountain+biking)true	Simple	I'd like to see a website about gibson guitars. website(i+d+like+to+see+a+website+about+gibson+guitars)false
	Simple	find me information about real dolls. links(information+about+real+dolls)true	Simple	find me a website about south park. website(south+park)true	*	I'd like to see a website about gibson guitars. website(i+d+like+to+see+a+website+about+gibson+guitars)false
	*	find me information about "real dolls". links(information+about+real+dolls)true	Simple	find me some scripts for blackadder. links(scripts+for+blackadder)true	Simple	I'd like to see a website about courtney love. website(i+d+like+to+see+a+website+about+courtney+love)false
	Simple	find me the atiitude clothing uk website. website(the+atiitude+clothing+uk)true		finshed at 1:53. (finshed+at+1:53)false	Simple	show me a picture of courtney love. picture(show+me+a+picture+of+courtney+love)false
	*	find me the atiitude clothing uk website. website(the+atiitude+clothing+uk)true			Simple	find me a website about queens of the stone age. website(queens+of+the+stone+age)true
	Simple	i want to buy something from attitude uk. (i+want+to+buy+something+from+attitude+uk)false			Simple	show me the fuct clothing website. website(show+me+the+fuct+clothing)true
	Simple	i want to buy something from attitude uk clothing. (i+want+to+buy+something+from+attitude+uk+clothing)false			*	show me the fuct clothing website. website(show+me+the+fuct+clothing)true
	Complex	find em the softback preview website, so i can reject this month's editors choice. website(find+em+the+softback+preview+website+so+i+can+reject+this+month+s+editors+choice)false			Simple	find me fuct clothing. links(fuct+clothing)true
	Simple	find me the softback preview website. website(the+softback+preview)true				
User 9	Question	what happened on the sopranos last night. (what+happened+on+the+sopranos+last+night)false	Keywords	radiohead. (radiohead)false	Complex	tell me about distributed information management. links(distributed+information+management)true
	Keywords	sopranos episode guide. (sopranos+episode+guide)false	Simple	tell me about the band called the sugarcubes. links(the+band+called+the+sugarcubes)true	Complex	what can you find on the subject of linking on the web. (what+can+find+on+the+subject+of+linking+on+the+web)false
	Simple	tell me about the channel 4 programme called bits. links(the+channel+4+programme+called+bits)true	Simple	find me some information on the rockingham speedway. links(information+on+the+rockingham+speedway)true	Complex	find me a paper on communicating using xml. links(a+paper+on+communicating+using+xml)true
	Simple	tell me about sony walkmans. links(sony+walkmans)true	Question	did anna kounikova win her last game?. (did+anna+kounikova+win+her+last+game)false	Question	who is working on exact tests for categorical data. (who+is+working+on+exact+tests+for+categorical+data)false
	Simple	tell me about canon. links(canon)true			Simple	find me something about go karting. links(something+about+go+karting)true
					Question	where is a good place to drink in southampton. (where+is+a+good+place+to+drink+in+southampton)false
User 10	Simple	I am interested in feature extraction from images.. pictures(i+am+interested+in+feature+extraction+from)true	Question	Are there any cycling clubs in the Bristol area?. (are+there+any+cycling+clubs+in+the+bristol+area)false	Question	Can you find me sites to do with knitting or cross-stich?. links(can+find+me+sites+to+do+with+knitting+or+cross-stich)false
	Question	Can you tell me about multimedia databases?. (can+tell+me+about+multimedia+databases)false	Question	Are there any cycling clubs in the Bristol (uk) area?. (are+there+any+cycling+clubs+in+the+bristol+(uk)+area)false	Question	Can you tell me whether Nick Cotton in Eastenders will ever walk again?. (can+tell+me+whether+nick+cotton+in+eastenders+ever+walk+again)false
	Question	I want to know what's happened to Nine Inch Nails. (i+want+to+know+what+s+happened+to+nine+inch+nails)false	*	Are there any cycling clubs in the Bristol (uk) area?. (are+there+any+cycling+clubs+in+the+bristol+(uk)+area)false	Simple	Show me sites about Eastenders. links(show+me+sites+about+eastenders)false
	Complex	Take me to the Nine Inch Nails and Trent Reznor web website(the+nine+inch+nails+and+trent+reznor+web+site)true	Question	Where can I buy metrology instruments for a cmm machine tool?. (where+can+i+buy+metrology+instruments+for+a+cmm+machine+tool)false		
			Question	Where can I get a Renishaw PH-10 CMM controller from?. (where+can+i+get+a+renishaw+ph-10+cmm+controller+from)false		
			Question	Where can I buy Renishaw equipment from?. (where+can+i+buy+renishaw+equipment+from)false		
User 11	Chitchat	West Ham United is the best football team ever. (west+ham+united+is+the+best+football+team+ever)false	Simple	find me stuff on the equator irc project. links(the+equator+irc+project)true	Insult	My god you're ugly. (my+god+re+ugly)false
	*	West Ham United is the best football team ever. (west+ham+united+is+the+best+football+team+ever)false	Insult	who inveted hypertext then you squared jawed surfer boy. (who+inveted+hypertext+then+squared+jawed+surfer+boy)false	Complex	tell me about adobe premiere books. links(adobe+premiere+books)true
	Question	who is number one in the world golf rankings. (who+is+number+one+in+the+world+golf+rankings)false	*	who invented hypertext then you squared jawed surfer boy. (who+invented+hypertext+then+squared+jawed+surfer+boy)false	Complex	find me the Billie holiday song What a little moonlight can do. links(the+billie+holiday+song+what+a+little+moonlight+can+do)true
	Chitchat	that anna komikova is a babe. (that+anna+komikova+is+a+babe)false	Question	have you even heard of Vannevar Bush?. (have+even+heard+of+vannevar+bush)false	Comples	what about dirty dude. (what+about+dirty+dude)false
	Simple	need a book on adobe premier. (need+a+book+on+adobe+premier)false	Question	give me some information on Monarch of the Glen, the BBC TV program. (give+me+information+on+monarch+of+the+glen+the+bbc+tv+program)false		
	NL	give me advice on directing a movie. (give+me+advice+on+directing+a+movie)false	Keywords	Monarch of the Glen, the BBC. (monarch+of+the+glen+the+bbc)false		

Keywords	27	Keywords	10	Keywords	17
Simple	13	Simple	9	Simple	10
Insult	0	Insult	1	Insult	1
Complex	2	Complex	3	Complex	7
Question	13	Question	14	Question	10
Chitchat	3	Chitchat	1	Chitchat	0
Rude	40	Rude	20	Rude	45
Polite	18	Polite	18	Polite	17
Rudeness Rating	0.69	Rudeness Rating	0.53	Rudeness Ratin	0.73
Politeness Rating	0.31	Politeness Rating	0.47	Politeness Ratin	0.27

Appendix 7

```
<html xmlns:v='urn:schemas-microsoft-com:vml'
xmlns:o='urn:schemas-microsoft-com:office:office'
xmlns:w='urn:schemas-microsoft-com:office:word'
xmlns='http://www.w3.org/TR/REC-html40'>


<meta http-equiv=Content-Type content='text/html; charset=windows-1252'>
<meta name=ProgId content=Word.Document>
<meta name=Generator content='Microsoft Word 9'>
<meta name=Originator content='Microsoft Word 9'>
<link rel=File-List href='./table_files/filelist.xml'>
<link rel=Edit-Time-Data href='./table_files/editdata.mso'>
<!--[if !mso]>
<style>
v\:* {behavior:url(#default#VML);}
o\:* {behavior:url(#default#VML);}
w\:* {behavior:url(#default#VML);}
.shape {behavior:url(#default#VML);}
</style>
<![endif]-->
<title> </title>
<!--[if gte mso 9]><xml>
 <o:DocumentProperties>
  <o:Author>Guillermo Power</o:Author>
  <o:Template>Normal</o:Template>
  <o:LastAuthor>Guillermo Power</o:LastAuthor>
  <o:Revision>2</o:Revision>
  <o:TotalTime>178</o:TotalTime>
  <o:Created>2002-07-18T17:38:00Z</o:Created>
  <o:LastSaved>2002-07-18T17:38:00Z</o:LastSaved>
  <o:Pages>1</o:Pages>
  <o:Company>University of Southampton</o:Company>
  <o:Lines>1</o:Lines>
  <o:Paragraphs>1</o:Paragraphs>
  <o:Version>9.2720</o:Version>
 </o:DocumentProperties>
</xml><![endif]-->
<style>
<!--
 /* Font Definitions */
@font-face
	{font-family:Verdana;
	panose-1:2 11 6 4 3 5 4 4 2 4;
	mso-font-charset:0;
	mso-generic-font-family:swiss;
	mso-font-pitch:variable;
	mso-font-signature:536871559 0 0 0 415 0;}
 /* Style Definitions */
p.MsoNormal, li.MsoNormal, div.MsoNormal
	{mso-style-parent:"";
	margin:0cm;
	margin-bottom:.0001pt;
	mso-pagination:widow-orphan;
	font-size:12.0pt;
	font-family:"Times New Roman";
	mso-fareast-font-family:"Times New Roman";}
p.MsoBodyText, li.MsoBodyText, div.MsoBodyText
	{margin:0cm;
	margin-bottom:.0001pt;
	text-align:center;
	mso-pagination:widow-orphan;
	font-size:26.0pt;
	mso-bidi-font-size:12.0pt;
	font-family:Verdana;
	mso-fareast-font-family:"Times New Roman";
	mso-bidi-font-family:"Times New Roman";}
@page Section1
	{size:595.3pt 841.9pt;
	margin:72.0pt 90.0pt 72.0pt 90.0pt;
	mso-header-margin:35.4pt;
	mso-footer-margin:35.4pt;
	mso-paper-source:0;}
div.Section1
	{page:Section1;}
-->
</style>


<meta http-equiv=Content-Type content='text/html; charset=windows-1252'>
<meta content=NOINDEX name=ROBOTS>
<meta content='MSHTML 6.00.2716.2200' name=GENERATOR></head>
<body bgColor=#ffffff topMargin=0><!--TOOLBAR_START--><!--TOOLBAR_EXEMPT--><!--TOOLBAR_END--><font
face=verdana,arial,helvetica size=2>
<center>


<hr width="66%">


<table border=0 cellspacing=0 cellpadding=0 style='border-collapse:collapse;
 mso-padding-alt:0cm 5.4pt 0cm 5.4pt'>
 <tr style='height:14.6pt'>
  <td width=190 rowspan=3 style='width:142.2pt;padding:0cm 5.4pt 0cm 5.4pt;
  height:14.6pt'>
  <p class=MsoNormal align=center style='text-align:center'><!--[if gte vml 1]><v:shapetype
   id="_x0000_t75" coordsize="21600,21600" o:spt="75" o:preferrelative="t"
   path="m@4@5l@4@11@9@11@9@5xe" filled="f" stroked="f">
   <v:stroke joinstyle="miter"/>
   <v:formulas>
    <v:f eqn="if lineDrawn pixelLineWidth 0"/>
    <v:f eqn="sum @0 1 0"/>
    <v:f eqn="sum 0 0 @1"/>
    <v:f eqn="prod @2 1 2"/>
    <v:f eqn="prod @3 21600 pixelWidth"/>
    <v:f eqn="prod @3 21600 pixelHeight"/>
    <v:f eqn="sum @0 0 1"/>
    <v:f eqn="prod @6 1 2"/>
    <v:f eqn="prod @7 21600 pixelWidth"/>
    <v:f eqn="sum @8 21600 0"/>
    <v:f eqn="prod @7 21600 pixelHeight"/>
    <v:f eqn="sum @10 21600 0"/>
   </v:formulas>
   <v:path o:extrusionok="f" gradientshapeok="t" o:connecttype="rect"/>
   <o:lock v:ext="edit" aspectratio="t"/>
  </v:shapetype><v:shape id="_x0000_i1072" type="#_x0000_t75" style='width:67.5pt;
   height:67.5pt'>
   <v:imagedata src="./table_files/image001.gif" o:title="southamptonLogo"/>
  </v:shape><![endif]--><![if !vml]><img width=90 height=90
  src="./table_files/image002.gif" v:shapes="_x0000_i1072"><![endif]></p>
  </td>
  <td width=378 style='width:283.45pt;padding:0cm 5.4pt 0cm 5.4pt;height:14.6pt'>
  <p class=MsoBodyText><b><span style='font-size:20.0pt;mso-bidi-font-size:
  12.0pt;font-family:"Arial Black"'>Southampton <o:p></o:p></span></b></p>

  </td>
  <td width=190 rowspan=3 style='width:142.25pt;padding:0cm 5.4pt 0cm 5.4pt;
  height:14.6pt'>
  <p class=MsoNormal align=center style='text-align:center'><!--[if gte vml 1]><v:shape
   id="_x0000_i1073" type="#_x0000_t75" style='width:51.75pt;height:63.25pt'>
   <v:imagedata src="./table_files/image003.gif" o:title="ipswichLogo"/>
```

```
   </v:shape><![endif]--><![if !vml]><img width=69 height=83
   src='./table_files/image004.gif' v:shapes='_x0000_i1073'><![endif]></p>
   </td>
  </tr>
  <tr style='height:14.6pt'>
   <td width=378 style='width:283.45pt;padding:0cm 5.4pt 0cm 5.4pt;height:14.6pt'>
   <p class=MsoBodyText><b><span style='font-size:20.0pt;mso-bidi-font-size:
   12.0pt;font-family:'Arial Black''>vs.<o:p></o:p></span></b></p>

   </td>
  </tr>
  <tr style='height:14.6pt'>
   <td width=378 style='width:283.45pt;padding:0cm 5.4pt 0cm 5.4pt;height:14.6pt'>
   <p class=MsoBodyText><b><span style='font-size:20.0pt;mso-bidi-font-size:
   12.0pt;font-family:'Arial Black''>Ipswich<o:p></o:p></span></b></p>
   </td>
  </tr>
 </table>


 <hr width='66%'>

 <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>

 <table border=0 cellspacing=0 cellpadding=0 style='border-collapse:collapse;
  mso-padding-alt:0cm 5.4pt 0cm 5.4pt'>
  <tr>
   <td width=191 valign=top style='width:143.6pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=108 valign=top style='width:81.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>

   <td width=180 valign=top style='width:135.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>

   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=96 valign=top style='width:72.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=180 valign=top style='width:135.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>

  </tr>
  <tr>
   <td width=191 valign=top style='width:143.6pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=108 valign=top style='width:81.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=180 valign=top style='width:135.0pt;padding:0cm 5.4pt 0cm 5.4pt'>

   <p class=MsoNormal align=center style='text-align:center'>Click <a
   href='http://stan/trial/question4.html'>HERE</a> when the animation finishes</p>
   </td>
   <td width=96 valign=top style='width:72.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=180 valign=top style='width:135.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>

   </td>
  </tr>
 </table>



</center><!--
In order to use Microsoft Agent, the Microsoft Agent Control
OBJECT tag must be placed on the page.  The presence of this
tag will cause the control to be automatically downloaded and
installed if it is not found on the client machine when the
page is processed.  In the example below, the CODEBASE
attribute is used so that the latest version number can be
specified.  The Agent object will be referred to in script using
the name assigned to it in the ID field of the OBJECT tag -
in this case, "AgentControl".
-->
<object id=AgentControl codeBase=#VERSION=2,0,0,0 height=0 width=0
classid=CLSID:D45FD31B-5C6E-11D1-9EC1-00C04FD7081F></object><!--
In order to use Text-to-Speech (TTS) output, a TTS engine compatible
with Microsoft Agent must be installed on the client's machine.
Your Microsoft Agent license includes a license to use the TruVoice
TTS engine from Lernout & Hauspie with Microsoft Agent.  The OBJECT tag
below causes the TTS engine to be downloaded and installed if it is
not found on the client machine when the page is processed. The
CODEBASE atribute is included in order to specify the latest version
number of the control.
-->
<object id=TruVoice codeBase=#VERSION=6,0,0,0 height=0 width=0
classid=CLSID:B8F2846E-CE36-11D0-AC83-00C04FD97575></object>
<script language=VBScript>
' In this example the Microsoft Agent control is managed using VBScript.
' Alternatively, JScript could be used.

Dim MyChar      ' a global variable to hold the character object

' Necessary initialization of the control and charcter are most
' readily accomplished in a page's OnLoad procedure, which is
' run automatically when the page is first loaded in a browser.
' The character to be used must first be loaded into the control.
' In this example the character is loaded from an HTTP URL, which
' must point to a .ACF file, and an object reference to the newly
' loaded character is saved in the global variable Robby.
' Once a character has been loaded from a .ACF
' file, it is necessary to GET each animation before PLAYing it.
' Here we GET all the animations that will be needed at once,
' rather than waiting until just before they are played.
' Newly loaded characters are initially hidden, so we show the
' character as the last step in initializing the character.  Since
' all the requests to the Agent control are queued and executed
' in request order, the appearance of the character will indicate
' that the preceding requests succeeded.  Note that this sample does
' not include any code to handle request failures.  A production
' quality page should always include error handling.
```

```
Sub window_OnLoad
	AgentControl.Connected = True
	AgentControl.Characters.Load "MyChar", "Emotion.acs"
	Set MyChar = AgentControl.Characters("MyChar")

	MyChar.LanguageID = &H0409		' needed under some conditions (English)
	' MyChar.Get "State", "Showing, Speaking"
	' MyChar.Get "Animation", "Greet, GreetReturn"
	MyChar.MoveTo 480,260
	MyChar.Show
	MyChar.Get "State", "Hiding"
	MyChar.Play "QuickSmile"

	MyChar.Speak "Welcome to Saints radio in a sunny Portland Road in this vital match in the Premiership. I think you can safely say that whoever wins today can breeth a lit

	MyChar.Speak "The players already on the pitch. The captains being called to the centre spot for the toss of the coin. Both Southampton and Eepswich will be going for a 

	MyChar.Speak "Long kick off, the ball is played back and Williams launches a long ball forwards. Davis gets his head to it and knocks it into the middle."

	MyChar.Speak "Knocked down by Marsden."

	MyChar.Speak "Goes to Brett Ormerod."

	MyChar.Speak "Ormerod leaves it to Paul Telfer."

	MyChar.Speak "Telfer drives it across."

	MyChar.Speak "Comes off the boot of Hermann Hreidarsson and within 20 seconds Southampton have won their first corner."

	' exhited
	MyChar.Play "shocked"
	MyChar.Speak "Taken quickly and short, it's tapped into the middle!!!!"
	MyChar.Play "Speaking"

	MyChar.Play "sad"
	MyChar.Speak "And it just goes over the head of Claus Lundekvam."
	MyChar.Speak "Marsden can only put it behind for a Goal kick to Eepswich town."
	MyChar.Play "Speaking"

	MyChar.Speak "The goal kick goes to Makin."

	MyChar.Speak "Makin finds Hredarsson."

	MyChar.Speak "Hredarsson dribbles it past LeTissier."

	MyChar.Speak "To Jamie Clappham."

	MyChar.Play "shocked"
	MyChar.Speak "Oh! Telfer steps in well and gets the ball. Intercepts."
	MyChar.Play "Speaking"

	MyChar.Speak "Telfer passes it back to Lundekvan."

	MyChar.Speak "Back to Telfer."

	MyChar.Speak "To LeTissier."

	MyChar.Speak "Southampton going forwards again."

	MyChar.Play "shocked"
	MyChar.Speak "LeTissier is still going!"
	MyChar.Speak "LeTissier finds Ormedon!"

	MyChar.Play "happy"
	MyChar.Speak "Ormedon scores!"

	MyChar.Play "angry"
	MyChar.Speak "No! The linesman raises his flag! Offside!!! I can't believe the linesman made a decision like that."

	MyChar.Play "sad"
	MyChar.Speak "We can see on the replay that Ormedon was well in. That's a very disappointing decision for Southampton"

	MyChar.Play "Speaking"
	MyChar.Speak "Long ball from Matteo."

	MyChar.Speak "Knocked down by Wright."

	MyChar.Play "scared"
	MyChar.Speak "Heavy weight collision there by Marcus Stewart and Fabrice Fernades."
	MyChar.Play "Speaking"

	MyChar.Speak "Southampton gets pocession of the ball."

	MyChar.Play "sad"
	MyChar.Speak "Fabrice Fernandes is still down. Ripley to LeTissier."
	MyChar.Play "Speaking"

	MyChar.Speak "LeTissier tries to go past Mark Venus. LeTissier's through."

	MyChar.Play "shocked"
	MyChar.Speak "Crosses it to Davies!!! Davies!!!!"
	MyChar.Play "Speaking"

	MyChar.Speak "Ooooh. Good goal keeping by Matteo"

	MyChar.Speak "That's got to go down as one of the saves of the season so far. Matteo really had to stretch for that one."

	MyChar.Speak "Southampton has been dominating so far."

	MyChar.Speak "Lundekvam to take the corner."

	MyChar.Speak "Lundekvam kicks."

	MyChar.Speak "A bit of a scramble and Wright kicks it away."

	MyChar.Speak "To Reuser."

	MyChar.Speak "Reuser under pressure from Fernandes."

	MyChar.Speak "Reuser keeps pocession and passes it to McGreal."

	MyChar.Speak "McGreal chips it back to Reuser."

	MyChar.Speak "Reuser gets it forward to Matt Holland."

	MyChar.Speak "Holland makes a run forwards and finds Bent."

	MyChar.Play "shocked"
	MyChar.Speak "Oooh!!!"
	MyChar.Play "Speaking"

	MyChar.Speak "Cheeky little chip from Bent and it goes over the bar. Goal kick."

	MyChar.Speak "That's Eepswich's best chance so far. The really caught the Southampton defence off guard there."

	MyChar.Speak "Jones takes the goal kick, chips it to Williams."

	MyChar.Speak "Williams turns it to Telfer, plays it forwards looking for the run from Mark Draper."

	MyChar.Play "shocked"
	MyChar.Speak "Draper crosses it!!!! Ormedon!!!! Brett Ormedon Scores!!!!!!"

	MyChar.Play "happy"
	MyChar.Speak "Lovely play that. That was a perfect cross from Draper."

	MyChar.Speak "That wraps that game up folks. It was bound to happen. Southampton dominated the game and there's only so much that Matteo can do to stop the Saint's inces
```

```
        Set Finale = MyChar.Hide

End Sub
</SCRIPT>
</font></body></html>
```

```html
<html xmlns:v='urn:schemas-microsoft-com:vml'
xmlns:o='urn:schemas-microsoft-com:office:office'
xmlns:w='urn:schemas-microsoft-com:office:word'
xmlns='http://www.w3.org/TR/REC-html40'>


<meta http-equiv=Content-Type content='text/html; charset=windows-1252'>
<meta name=ProgId content=Word.Document>
<meta name=Generator content='Microsoft Word 9'>
<meta name=Originator content='Microsoft Word 9'>
<link rel=File-List href='./table_files/filelist.xml'>
<link rel=Edit-Time-Data href='./table_files/editdata.mso'>
<!--[if !mso]>
<style>
v\:* {behavior:url(#default#VML);}
o\:* {behavior:url(#default#VML);}
w\:* {behavior:url(#default#VML);}
.shape {behavior:url(#default#VML);}
</style>
<![endif]-->
<title> </title>
<!--[if gte mso 9]><xml>
 <o:DocumentProperties>
  <o:Author>Guillermo Power</o:Author>
  <o:Template>Normal</o:Template>
  <o:LastAuthor>Guillermo Power</o:LastAuthor>
  <o:Revision>2</o:Revision>
  <o:TotalTime>178</o:TotalTime>
  <o:Created>2002-07-18T17:38:00Z</o:Created>
  <o:LastSaved>2002-07-18T17:38:00Z</o:LastSaved>
  <o:Pages>1</o:Pages>
  <o:Company>University of Southampton</o:Company>
  <o:Lines>1</o:Lines>
  <o:Paragraphs>1</o:Paragraphs>
  <o:Version>9.2720</o:Version>
 </o:DocumentProperties>
</xml><![endif]-->
<style>
<!--
 /* Font Definitions */
@font-face
	{font-family:Verdana;
	panose-1:2 11 6 4 3 5 4 4 2 4;
	mso-font-charset:0;
	mso-generic-font-family:swiss;
	mso-font-pitch:variable;
	mso-font-signature:536871559 0 0 0 415 0;}
 /* Style Definitions */
p.MsoNormal, li.MsoNormal, div.MsoNormal
	{mso-style-parent:"";
	margin:0cm;
	margin-bottom:.0001pt;
	mso-pagination:widow-orphan;
	font-size:12.0pt;
	font-family:"Times New Roman";
	mso-fareast-font-family:"Times New Roman";}
p.MsoBodyText, li.MsoBodyText, div.MsoBodyText
	{margin:0cm;
	margin-bottom:.0001pt;
	text-align:center;
	mso-pagination:widow-orphan;
	font-size:26.0pt;
	mso-bidi-font-size:12.0pt;
	font-family:Verdana;
	mso-fareast-font-family:"Times New Roman";
	mso-bidi-font-family:"Times New Roman";}
@page Section1
	{size:595.3pt 841.9pt;
	margin:72.0pt 90.0pt 72.0pt 90.0pt;
	mso-header-margin:35.4pt;
	mso-footer-margin:35.4pt;
	mso-paper-source:0;}
div.Section1
	{page:Section1;}
-->
</style>


<meta http-equiv=Content-Type content='text/html; charset=windows-1252'>
<meta content=NOINDEX name=ROBOTS>
<meta content='MSHTML 6.00.2716.2200' name=GENERATOR></head>
<body bgColor=#ffffff topMargin=0><!--TOOLBAR_START--><!--TOOLBAR_EXEMPT--><!--TOOLBAR_END--><font
face=verdana,arial,helvetica size=2>
<center>


<hr width='66%'>


<table border=0 cellspacing=0 cellpadding=0 style='border-collapse:collapse;
 mso-padding-alt:0cm 5.4pt 0cm 5.4pt'>
 <tr style='height:14.6pt'>
  <td width=190 rowspan=3 style='width:142.2pt;padding:0cm 5.4pt 0cm 5.4pt;
  height:14.6pt'>
  <p class=MsoNormal align=center style='text-align:center'><!--[if gte vml 1]><v:shapetype
   id="_x0000_t75" coordsize="21600,21600" o:spt="75" o:preferrelative="t"
   path="m@4@5l@4@11@9@11@9@5xe" filled="f" stroked="f">
   <v:stroke joinstyle="miter"/>
   <v:formulas>
    <v:f eqn="if lineDrawn pixelLineWidth 0"/>
    <v:f eqn="sum @0 1 0"/>
    <v:f eqn="sum 0 0 @1"/>
    <v:f eqn="prod @2 1 2"/>
    <v:f eqn="prod @3 21600 pixelWidth"/>
    <v:f eqn="prod @3 21600 pixelHeight"/>
    <v:f eqn="sum @0 0 1"/>
    <v:f eqn="prod @6 1 2"/>
    <v:f eqn="prod @7 21600 pixelWidth"/>
    <v:f eqn="sum @8 21600 0"/>
    <v:f eqn="prod @7 21600 pixelHeight"/>
    <v:f eqn="sum @10 21600 0"/>
   </v:formulas>
   <v:path o:extrusionok="f" gradientshapeok="t" o:connecttype="rect"/>
   <o:lock v:ext="edit" aspectratio="t"/>
  </v:shapetype><v:shape id="_x0000_i1072" type="#_x0000_t75" style='width:67.5pt;
   height:67.5pt'>
   <v:imagedata src="./table_files/image001.gif" o:title="southamptonLogo"/>
  </v:shape><![endif]--><![if !vml]><img width=90 height=90
  src="./table_files/image002.gif" v:shapes='_x0000_i1072'><![endif]></p>
  </td>
  <td width=378 style='width:283.45pt;padding:0cm 5.4pt 0cm 5.4pt;height:14.6pt'>
  <p class=MsoBodyText><b><span style='font-size:20.0pt;mso-bidi-font-size:
  12.0pt;font-family:"Arial Black"'>Southampton <o:p></o:p></span></b></p>

  </td>
  <td width=190 rowspan=3 style='width:142.25pt;padding:0cm 5.4pt 0cm 5.4pt;
  height:14.6pt'>
  <p class=MsoNormal align=center style='text-align:center'><!--[if gte vml 1]><v:shape
   id="_x0000_i1073" type="#_x0000_t75" style='width:51.75pt;height:63.25pt'>
   <v:imagedata src="./table_files/image003.gif" o:title="ipswichLogo"/>
```

```
   </v:shape><![endif]--><![if !vml]><img width=69 height=83
   src='./table_files/image004.gif' v:shapes='_x0000_i1073'><![endif]></p>
   </td>
  </tr>
  <tr style='height:14.6pt'>
   <td width=378 style='width:283.45pt;padding:0cm 5.4pt 0cm 5.4pt;height:14.6pt'>
   <p class=MsoBodyText><b><span style='font-size:20.0pt;mso-bidi-font-size:
   12.0pt;font-family:'Arial Black''>vs.<o:p></o:p></span></b></p>

   </td>
  </tr>
  <tr style='height:14.6pt'>
   <td width=378 style='width:283.45pt;padding:0cm 5.4pt 0cm 5.4pt;height:14.6pt'>
   <p class=MsoBodyText><b><span style='font-size:20.0pt;mso-bidi-font-size:
   12.0pt;font-family:'Arial Black''>Ipswich<o:p></o:p></span></b></p>
   </td>
  </tr>
 </table>


 <hr width="66%">

 <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>

 <table border=0 cellspacing=0 cellpadding=0 style='border-collapse:collapse;
  mso-padding-alt:0cm 5.4pt 0cm 5.4pt'>
  <tr>
   <td width=191 valign=top style='width:143.6pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=108 valign=top style='width:81.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>

   <td width=180 valign=top style='width:135.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>

   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=96 valign=top style='width:72.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=180 valign=top style='width:135.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>

  </tr>
  <tr>
   <td width=191 valign=top style='width:143.6pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=108 valign=top style='width:81.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=180 valign=top style='width:135.0pt;padding:0cm 5.4pt 0cm 5.4pt'>

   <p class=MsoNormal align=center style='text-align:center'>Click <a
   href="http://stan/trial/question3.html">HERE</a> when the animation finishes</p>
   </td>
   <td width=96 valign=top style='width:72.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>
   </td>
   <td width=180 valign=top style='width:135.0pt;padding:0cm 5.4pt 0cm 5.4pt'>
   <p class=MsoNormal><![if !supportEmptyParas]> <![endif]><o:p></o:p></p>

   </td>
  </tr>
 </table>



</center><!--
In order to use Microsoft Agent, the Microsoft Agent Control
OBJECT tag must be placed on the page.  The presence of this
tag will cause the control to be automatically downloaded and
installed if it is not found on the client machine when the
page is processed.  In the example below, the CODEBASE
attribute is used so that the latest version number can be
specified.  The Agent object will be referred to in script using
the name assigned to it in the ID field of the OBJECT tag -
in this case, "AgentControl".
-->
<object id=AgentControl codeBase=#VERSION=2,0,0,0 height=0 width=0
classid=CLSID:D45FD31B-5C6E-11D1-9EC1-00C04FD7081F></object><!--
In order to use Text-to-Speech (TTS) output, a TTS engine compatible
with Microsoft Agent must be installed on the client's machine.
Your Microsoft Agent license includes a license to use the TruVoice
TTS engine from Lernout & Hauspie with Microsoft Agent.  The OBJECT tag
below causes the TTS engine to be downloaded and installed if it is
not found on the client machine when the page is processed. The
CODEBASE atribute is included in order to specify the latest version
number of the control.
-->
<object id=TruVoice codeBase=#VERSION=6,0,0,0 height=0 width=0
classid=CLSID:B8F2846E-CE36-11D0-AC83-00C04FD97575></object>
<script language=VBScript>
' In this example the Microsoft Agent control is managed using VBScript.
' Alternatively, JScript could be used.

Dim MyChar      ' a global variable to hold the character object

' Necessary initialization of the control and charcter are most
' readily accomplished in a page's OnLoad procedure, which is
' run automatically when the page is first loaded in a browser.
' The character to be used must first be loaded into the control.
' In this example the character is loaded from an HTTP URL, which
' must point to a .ACF file, and an object reference to the newly
' loaded character is saved in the global variable Robby.
' Once a character has been loaded from a .ACF
' file, it is necessary to GET each animation before PLAYing it.
' Here we GET all the animations that will be needed at once,
' rather than waiting until just before they are played.
' Newly loaded characters are initially hidden, so we show the
' character as the last step in initializing the character.  Since
' all the requests to the Agent control are queued and executed
' in request order, the appearance of the character will indicate
' that the preceding requests succeeded.  Note that this sample does
' not include any code to handle request failures.  A production
' quality page should always include error handling.
```

```
Sub window_OnLoad
        AgentControl.Connected = True
        AgentControl.Characters.Load "MyChar", "Emotion.acs"
        Set MyChar = AgentControl.Characters("MyChar")

        MyChar.LanguageID = &H0409              ' needed under some conditions (English)
        ' MyChar.Get "State", "Showing, Speaking"
        ' MyChar.Get "Animation", "Greet, GreetReturn"
        MyChar.MoveTo 480,260
        MyChar.Show
        MyChar.Get "State", "Hiding"

        MyChar.Speak "Welcome to Saints radio in a sunny Portland Road in this vital match in the Premiership. I think you can safely say that whoever wins today can breeth a li

        MyChar.Speak "The players already on the pitch. The captains being called to the centre spot for the toss of the coin. Both Southampton and Eepswich will be going for a

        MyChar.Speak "Long kick off, the ball is played back and Williams launches a long ball forwards. Davis gets his head to it and knocks it into the middle."

        MyChar.Speak "Knocked down by Marsden."

        MyChar.Speak "Goes to Brett Ormerod."

        MyChar.Speak "Ormerod leaves it to Paul Telfer."

        MyChar.Speak "Telfer drives it across."

        MyChar.Speak "Comes off the boot of Hermann Hreidarsson and within 20 seconds Southampton have won their first corner."

        MyChar.Speak "Taken quickly and short, it's tapped into the middle!!!!"

        MyChar.Speak "And it just goes over the head of Claus Lundekvam."
        MyChar.Speak "Marsden can only put it behind for a Goal kick to Eepswich town."

        MyChar.Speak "The goal kick goes to Makin."

        MyChar.Speak "Makin finds Hredarsson."

        MyChar.Speak "Hredarsson dribbles it past LeTissier."

        MyChar.Speak "To Jamie Clappham."

        MyChar.Speak "Oh! Telfer steps in well and gets the ball. Intercepts."

        MyChar.Speak "Telfer passes it back to Lundekvan."

        MyChar.Speak "Back to Telfer."

        MyChar.Speak "To LeTissier."

        MyChar.Speak "Southampton going forwards again."

        MyChar.Speak "LeTissier is still going!"
        MyChar.Speak "LeTissier finds Ormedon!"

        MyChar.Speak "Ormedon scores!"

        MyChar.Speak "No! The linesman raises his flag! Offside!!! I can't believe the linesman made a decision like that."

        MyChar.Speak "We can see on the replay that Ormedon was well in. That's a very disappointing decision for Southampton"

        MyChar.Speak "Long ball from Matteo."

        MyChar.Speak "Knocked down by Wright."

        MyChar.Speak "Heavy weight collision there by Marcus Stewart and Fabrice Fernades."

        MyChar.Speak "Southampton gets pocession of the ball."

        MyChar.Speak "Fabrice Fernandes is still down. Ripley to LeTissier."

        MyChar.Speak "LeTissier tries to go past Mark Venus. LeTissier's through."

        MyChar.Speak "Crosses it to Davies!!! Davies!!!!"

        MyChar.Speak "Ooooh. Good goal keeping by Matteo"

        MyChar.Speak "That's got to go down as one of the saves of the season so far. Matteo really had to stretch for that one."

        MyChar.Speak "Southampton has been dominating so far."

        MyChar.Speak "Lundekvam to take the corner."

        MyChar.Speak "Lundekvam kicks."

        MyChar.Speak "A bit of a scramble and Wright kicks it away."

        MyChar.Speak "To Reuser."

        MyChar.Speak "Reuser under pressure from Fernandes."

        MyChar.Speak "Reuser keeps pocession and passes it to McGreal."

        MyChar.Speak "McGreal chips it back to Reuser."

        MyChar.Speak "Reuser gets it forward to Matt Holland."

        MyChar.Speak "Holland makes a run forwards and finds Bent."

        MyChar.Speak "Oooh!!!"

        MyChar.Speak "Cheeky little chip from Bent and it goes over the bar. Goal kick."

        MyChar.Speak "That's Eepswich's best chance so far. The really caught the Southampton defence off guard there."

        MyChar.Speak "Jones takes the goal kick, chips it to Williams."

        MyChar.Speak "Williams turns it to Telfer, plays it forwards looking for the run from Mark Draper."

        MyChar.Speak "Draper crosses it!!!! Ormedon!!!! Brett Ormedon Scores!!!!!!"

        MyChar.Speak "Lovely play that. That was a perfect cross from Draper."

        MyChar.Speak "That wraps that game up folks. It was bound to happen, Southampton dominated the game and there's only so much that Matteo can do to stop the Saint's inces

        Set Finale = MyChar.Hide

End Sub
</SCRIPT>
```