Metaphor to Personality : the role of animation in intelligent interface agents

Michael Wilson
Rutherford Appleton Laboratory
Chilton
Didcot
Oxon OX11 0QX
UK
mdw@inf.rl.ac.uk

Abstract

Intelligent agents are beginning to appear that incorporate animated user interfaces. This paper derives 10 requirements on the interfaces to such agents in the light of the role the agent takes with respect to the user, the mental model the user is expected to develop of the agent, and the aspects of processes which are captured by animation. It is concluded that the main role of the animated interface is to support the marketing of the interface through the creation of an acceptable characterisation of it, rather than to support interactive use of the agent.

1 Introduction

Intelligent agents are beginning to appear which are presented to users as animated characters. This paper analyses the requirements on the interfaces to such agents. Following these requirements the interface of one animated agent system is analysed to answer three questions which should guide the designers of animated agents :

Does the introduction of character and the animation of representations of those characters for intelligent agents contribute to meeting these requirements ?

If so, what are the dimensions in the design space of these characters and animated representations that designers can manipulate ?

Thirdly, what aspects of the task are not performed by the characters and their animated representations which must be met otherwise, and what constraints do the animated character representations place on these ancillary mechanisms ?

The second section presents the established HCI view on the mental models users develop of systems, and the interaction styles available at the user interface to do this. The third section then looks at the mental models which users need to develop in order to interact with intelligent agents successfully and the requirements these place on the design of agent representations. The fourth section then investigates the presentation of active processes to users, using simple semiotic distinctions to provide further requirements on the design of animated representations of intelligent agents. Section five then analyses an existing character based animated interface to an intelligent agent in terms of these requirements, and the principles behind the craft skill of film animation. This leads into a discussion of the reality of the representation as a metaphor for the intelligent agent and the complexity of interaction required by human like representations.

2 User Mental Models and Interaction Styles

Traditional HCI design is aimed at hiding the implementation model of the computer process from users and developing a clear mental model for the user of their interaction with the computer in order to complete their task goals (e.g. Newman and Lamming, 1995). Four major classes of mental model are available and interface designers choose which one they wish the user to hold and design the interface to support it :

Object Action Model - The user is aware of the objects in the system and the actions which can be performed on those objects. Desktop objects which can be labeled, moved, opened or destroyed use this class of model.

State Transition Model - systems which are moded and change modes are usually presented through state transition models; for example, telephone systems were designed so that the state transitions were audible to users through dial tones.

Mapping Model - when an intention is mapped into a sequence of actions, normally a repeated sequence of actions have to be performed. For example, air-line check-in clerks map the intention of querying the system about a passenger to the command line sequence to be keyed in. The mapping model takes over from the state transition or object-action models when the sequence of actions is repeated so as to become proceduralised at the execution level.

Analogical Model - metaphor or analogical models apply when users encounter a system which closely resembles one that they are familiar with. The desktop analogy is of this class, where the analogy serves to introduce users to concepts in the user interface. Analogy models can be of any of the other types above.

Users may change the class of model from being metaphor through state-transition or object-action model to mapping as they become familiar with a system. They may also jump between models for different parts of a system, a they are more experienced with some parts that are proceduralised to support the mapping model, while requiring an analogical model for other parts with which they have less or no experience.

The interface design options available to a designer fall broadly into classes of interaction style which can be mapped onto the desired mental model :

Key Modal

Menu-based Interaction (including hypertext)
Question and Answer
Function Key Interaction
Voice-based Interaction

Almost all walk-up-and-use systems are key-modal. State changes or mode changes occur as a result of function key or mouse selections. These follow a state transition model. They are limited in expression, but good for navigation where context provides the meaning (by presenting a set of contrasting alternatives from which the semantics of each choice can be derived, or through embedding options in longer explanations).

Direct Manipulation

Graphical Direct Input
Forms Fill In
Virtual Reality

Direct manipulation uses an object-action model, and is usually combined with a metaphor as a basis for the interpretation of the semantics of the object and its accompanying actions.

Linguistic

Command Language
Textual Natural Language

The linguistic style supports the expression of complex commands and requests such as information retrieval or database searches. Command languages require a long learning time compared to other styles since their is no support for either terms, or their semantics during use, so that they rely on human memory retrieval rather than recognition. Languages are often designed to be analogs of known languages, although the transfer of semantics usually fails as in the case of the relation between the Boolean AND and OR and those terms in natural language. Natural Language interaction assumes that users already know the language which is unfortunately an error since human natural languages are really a collection of sub-languages spoken by different individuals or groups in different situations and not a single language (despite Chomsky's competence theory of language). Human natural language is supported by facilities in dialogue to negotiate the meaning and use of terms, and to clarify and disambiguate meaning which consume a large proportion of normal human conversation. Unless these mechanisms exist for human computer interaction, the human language metaphor for human computer interaction breaks down.

In practice designers combine these different styles but aim to ensure that the underlying mental models developed by users change consistently with the style changes.

3 The Mental Model for Intelligent Assistants

Experience from the development of Expert Systems in the 1980's (e.g. Boy, 1991) suggested that the relationship between the user and an intelligent process would place the process in one of the following four roles :

Tool - the process does what the user tells it. Classic computer processes are designed as tools, which are instruments with which the user acts upon data, or which are used to control remote devices. Each tool has clearly defined inputs and outputs, and the user can expect that the tool will perform as required as long as the input is correct (assuming that resources are available for its functioning). The output will either be that expected from the successful execution of the tool, or an error message detailing mistakes or slips in the input. Each of the mental models described above, with the corresponding interaction styles could be used in designing interfaces to tools.

Guru - the process performs some reasoning and tells the user the results. This is typical of medical diagnostic programs where the user describes the symptoms, the system performs the diagnosis and informs the user who then undertakes treatment. One of the main problems with classic diagnostic medical expert systems was that the users were expected to be highly prestigious medical doctors who were neither satisfied with following the guidance of a process, nor willing to accept the legal position which following its advice entailed. A consequence of this was that many systems developed explanation facilities to move their role to that of colleague.

Colleague - the user and process negotiate about the issue of consideration in a way akin to asking a colleague for a second opinion, and negotiating if it is at variance with one's own. For example, a medical practitioner may describe a patients symptoms to a system along with her diagnosis and planned course of treatment; the process would then critique the practitioner's diagnostic and planning reasoning in order to identify possible errors, confounds, or to suggest likely alternatives to ensure that the practitioner had considered them and ruled them out. As with colleagues, one can ignore the advice of the process, and if the process consistently gives advice which is ignored because it is erroneous itself, or is not sufficiently justified through negotiation to convince the user, the process itself will cease to be consulted.

Assistant - Operators who control complex dynamic systems, generally use mechanical or human aids, e.g. operation manuals or users guides. In addition to these aids, there is often assistance from other human operators called assistants, copilots or collaborators. For example, in an aircraft cockpit, a human copilot shares the work, but not the ultimate responsibility, with the captain. If the captain delegates a part of his responsibility to the copilot, then the copilot will take this delegation as a task to be executed. In addition, the captain may choose at any time to stop the execution of a task by the copilot. However, a copilot may have personal initiatives, for example, testing parameters, keeping current with the evolving situation, predicting deducible faults, etc. He should be capable of explaining, in an appropriate amount of detail, the results of his processing.

Intelligent agents with user interfaces can be capable of playing any of the last three roles, although the last role of assistant is the one which is currently most visible in applications such as Mail sorting, calendar and meeting organising, WWW information locating, music and entertainment selection (e.g. Maes, 1994). The mental model that should be created in the user to support this is exemplified by the above description, but it must avoid confusion with the models of the other three categories. Obviously, within this general mental model, more specific analogies with particular assistants beyond airline copilots can be developed to show the range of knowledge or skills held by the assistant, but it must be clear to the user what the boundaries of these are, and where responsibilities as well as tasks can be delegated.

The metaphor generally used for expert systems was that of an intelligent human being, expert in one area. Unfortunately users over-generalized this intelligence since they had no mental model that disassociated specialist reasoning from general reasoning skills, or specialist knowledge from general knowledge and general communication skills. Consequently, the localised knowledge and brittle reasoning within the limited scope of a knowledge base even when explained through various dialogue techniques could never establish a clear mental model which exactly characterised the scope of the skills and knowledge of most expert systems with the result that users frequently crossed over the cliff at the edge o the plateau of their abilities, and subsequently could not trust them, even to do what they were able to.

Beyond expressing the range of task relevant knowledge and skills possessed by the intelligent assistant, the assistant role requires the user interface to support the following activities by the user :

1) Delegate tasks and responsibilities to the agent - the interface must afford user actions to do this.
2) Evaluate the execution of a task in order to decide to terminate the delegation.
3) Explain the results of its reasoning., and from that explanation, decide on the competence of the agent sufficiently to assess whether to delegate further tasks.
4) To promote a mental model in the user consistent with the abilities of the assistant and compatible with the user's representation of the task.

4 Presenting Active Processes

Psychological theory such as that at the foundation of HCI has little detail to add about film and animation specifically. Less objective analyses of this area are provided within the framework of semiotics. Semiotic theory (e.g. Stam et al, 1992) derived from the philosophy of Charles Pierce (1839-1914) defines signs as "something which stands to somebody for something in some respects or capacity". The process of production of meaning involves a triad of three entities: the sign, its object (that for which the sign stands) and its interpretant (that is the mental effect generated by the relation between the sign and object). Signs are further subdivided into icons, which represent an object by virtue of some aspect of their internal nature (e.g. a portrait, statue, onomatopoeic word); indices which involves a causal link between the sign and interpretant (e.g. a weather cock or a barometer); and symbols which involve entirely conventional links between sign and interpretant (e.g. linguistic signs are symbols in that they represent objects only by linguistic convention). Before progressing with semiotic theory, we will analyse the presentation of processes of various complexity within these initial semiotic distinctions where the designer is intending to develop a user's mental model through metaphors to a state transition model.

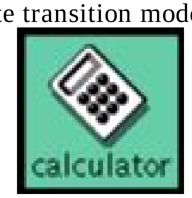

Figure 1: Calculator desktop representation from the SunView user interface (circa 1984) presenting both a pictorial icon and a textual symbol to disambiguate it.

The simple state transition model assumed is one where *processes* are divided into *states* which are linked by *events*, and sets of states and linking events can be hidden inside black boxes from the user as *indivisible processing states* themselves which take *time* to perform. The simplest process involves two states (active and inactive) with an activation or deactivation events to move between them. An example is shown in Figure 1, of a calculator desktop representation from the Sunview user interface (circa 1984). This presents both a pictorial icon and a textual symbol to disambiguate it in case it is either ambiguous itself, or is confusable with other images available to the user. The icon serves the purpose of stating the existence of the tool, and the metaphor it subsumes of being a desktop button affords its selection with a mouse in order to activate it. Once active it can be closed down to its current closed state. The open state itself is presented as a metaphor from hand held calculators, with buttons, but further analysis of the open state does not further the view of the static closed icon. It is important to note that the icon represents existence of the tool where users may not know that the tool is available in which case they may be searching for a representation of an object to fulfill the desired functionality to perform their task, therefore the representation conveys the functionality through the icon. Once the icon is identified it may become conventionalised by users who recall it as a calculator rather than relying on the iconic resemblance. It is also worth noting that the icon is designed as part of an open system, where the designer does not know the contrast set of icons or the set of tasks which users may be performing, therefore the contrast between this image and others available to users cannot be known to the designer. Consequently, ambiguities and confusions are possible and the textual symbol is used to reduce these.

Figure 2: Animated mail tool representation from the SunView user interface presenting a metaphor of a mail in-tray in three states as an index with textual symbol: empty, contains new mail, contains read mail.

The next stage of complexity is to represent the states of an active process as an index. A simple example of this is shown in Figure 2, also from the SunView system, where a tool for handling electronic mail can hold any of three states (empty, contains new mail, contains read mail) and the index is animated to capture the dynamic changes in state. The design issues from the icon remain concerning ambiguity, confusability and affordance to action, and the same solutions have been applied.

In the two previous examples activity was limited to instant events that caused state changes, the next addition to this to represent the whole of the simple state transition model is to animate an index representation where activity within a process which is not subdivided into states and events is presented as taking place over time. Such representations reassure users that some activity is taking place and that a global failure has not occurred, and if they are active, they also fill time, rather than leaving empty unfilled time to depress the user. For single tasking operating systems this was normally represented by changing the cursor to an icon representing time (e.g. an hourglass) since the cursor was the focus of user attention. In multitasking OS this approach can be used when the cursor is held over a screen area (e.g. window) representing a task, but this requires user action to follow their change in attention focus to that area, away from their current task. An example of a display which does not require action to move away from the current focus of attention is shown in Figure 3 as an animated file download dialogue from MS-Windows 95/NT. This presents a metaphor of sheets of paper floating from one folder to another as a file is transferred in order to represent the activity of the process. As in the previous examples, the index is supplemented by symbols to disambiguate, and clarify the process, even in such a simple example. This index is supplemented by a textual symbol banner to disambiguate the activity presented by the index, and by a description of the file to disambiguate the object involved in the operation in order to disambiguate this from any other file download process. As in the earlier example, only one action is afforded by the presentation, but here it is a cancellation of the active process, rather than an opening of a tool. This action is afforded by an icon of a button which is labeled by a textual symbol to disambiguate its function.

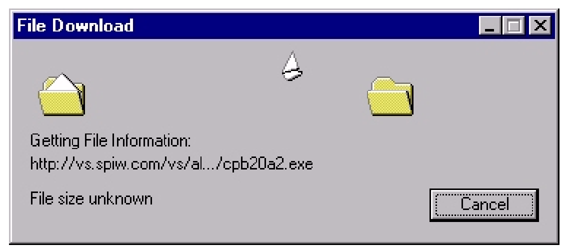

Figure 3: Animated file download dialogue from MS-Windows 95/NT presenting a metaphor of sheets of paper floating from one folder to another as a file is transferred in order to represent the activity of the process

Figure 4: Animated file download dialogue from MS-Windows 95/NT with the addition of the amount of the process completed, and the expected duration of the remainder of the process.

The previous example presented an indivisible process as an active state. Indeed in that example A textual symbol was added to the presentation to state the stage that process held in a larger divisible process (“Getting File Information”). Two further additions can be made to the semantics expressed to an active indivisible process which is to state the amount of the process completed, and the expected duration of the remainder of the process. The example in Figure 4 shows a later stage in the same File Download where the next indivisible process has been entered (Saving) and this has both of these items of information; the expected duration of the remainder of the process added to it as a textual symbol, and the amount of the process completed shown through a meter metaphor, adding a second animated element to the overall compound index & symbol set.

This laboured analysis using the simplest semiotic elements of icons, indexes and symbols in user interface dialogue has shown that many mechanisms are required to avoid the disambiguations and clarifications that human dialogue supports in a dialogue of even the slightest complexity. These examples of animated indexes of processes as compounds of indexes & symbols have been used to establish mental models built from state transition models and analogical models need to present the following six elements, which become further requirements on any animation of an intelligent assistant:

5) icons to convey existence and metaphoric semantics of state
6) indexes to convey events and state change
7) indexes to convey indivisible process activity
indexes to convey indivisible process duration
symbols to disambiguate and clarify the other signs
affordance of user actions available.

5 Intelligent Assistants at the Interface

For complex programs, many programmers produce small animations of high level abstractions of a state transition diagram to aid debugging, by showing that the program is still active, which high level component is currently running, and whether the program has not entered an infinite loop. These animated state transition diagrams are only intended for programmers who already know the structure of the program and can relate the structure of the implementation model to the behaviour of the program. These do not attempt to show what actions can be taken (usually abort and debug when a loop is found) nor relate the implementation model to a metaphorical assistant model in order that the behaviour can be related to a user's task performance.

It appears that the simplest form of animation at the user interface (e.g. .Maes, 1994) is a development of this debugging tool where the states of high level units in the implementation model are characterised as states of an anthropomorphised assistant undertaking the task. This simple mapping from implementation states to states of human task behaviour are obviously derived from the simple metaphor of the human assistant. However, whether it meets all of the ten requirements so far established requires a little more analysis.

The Autonomy agent (Autonomy Corporation, 1996) is a simple WWW information finding agent, which requires training on the class of information to be located (as shown in figure 7), then searches the web for pages which contain that information, retrieves relevant information, and notifies the user of its success. After inspection of the retrieved information the user may wish to refine or expand the search criteria by re-training the agent using either search terms or by marking retrieved items as incorrect. The Autonomy agent index illustrated in figure 5 is like the mailtool index described above on that it appears to meet requirements 1 and 7 listed in the previous section and not meet requirements 2-6 or 8-10.

This example includes 10 states of the index which are animated, but when the agent is either active, or inactive the index is still animated since the character "shuffles around impatiently". The icon of a dog clearly conveys a metaphor of a loyal, faithful, dumb agent which will obey his master's instructions most of the time, but may playfully follow his own volition at other times. Firstly, the dog connotes some of these responses itself, and then the representation chosen uses visual indicators of character psychology such as large eyes and the facial proportions of a young child rather than an adult. This metaphor avoids the problems of *guru* expert systems in that the model is not one from which users would take instruction, nor one to which they would overly delegate authority. Therefore the animation superficially addresses Requirement 1. Indeed this image is comforting, amusing and even lovable. In general this image supports the marketing of the product, but does it add anything by the animated index of the system's states ?

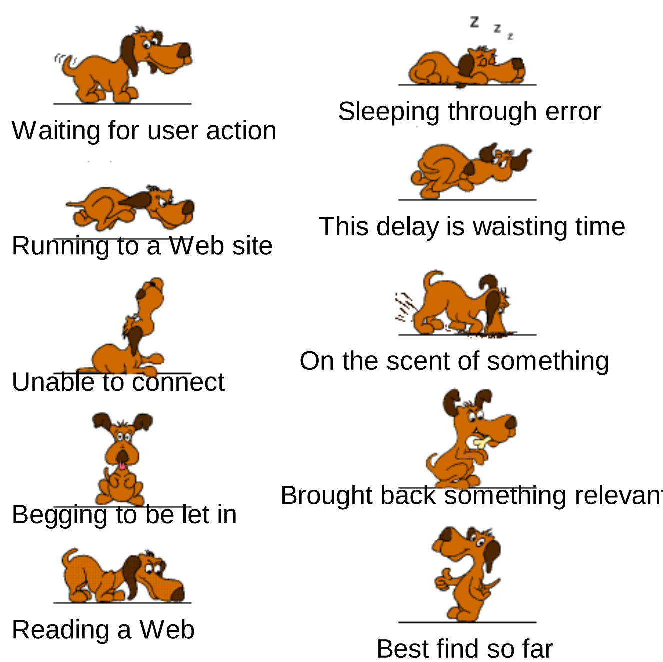

Figure 5: The ten states of the animated index for an agent in the Autonomy Agent system for locating information on the WWW.

It clearly shows that some state is changing, but the text in Figure 5 is added into the figure as a legend and does not occur as a textual symbol on the index. Without this text does the image convey the state that is being represented ? The analogy of a dog begging to be let in, digging for a bone and returning with a bone in its mouth are not unreasonable analogies to search and retrieval. But the other images seem forced within this metaphor, and require the text to explain them as in Requirement 9. Therefore, in use, the actual index conveys that the system is active because of the animation (Requirement 7), but does not inform the user of what it is doing so that the user cannot evaluate the task performance in accordance with Requirement 2 above.

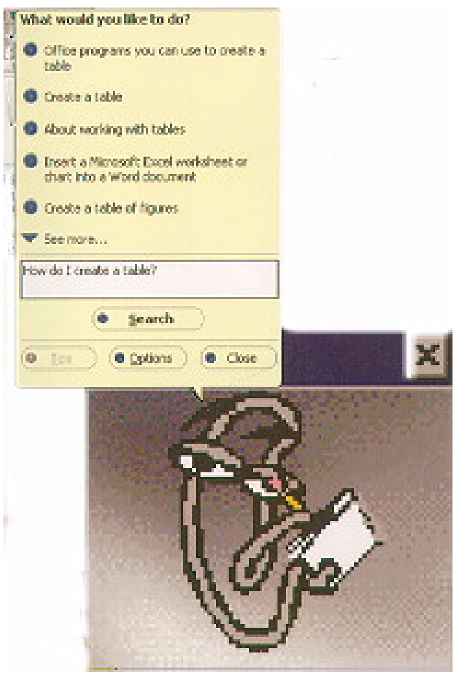

Figure 6: The paperclip assistant agent from MS-Office '97 uses pop-up dialogue boxes.

The index clearly does not indicate the duration of the task ahead, or the amount completed (requirement 8), and it equally indicates nothing about the actions or communication that the user can make to control it (requirement 10). The Autonomy agent uses conventional typed dialogue to train it and report (though not explain) its results as shown in Figure 7. The metaphor of the trained hound is continued through these interactions graphically, although they are not animated and do not add to the establishment of a complete mental model of the assistant or its interaction. The change to the combined natural language and button based interface provides a break in interaction style consistent with the break in the mental model, so that a model where the user is programming through natural language can be developed. The interaction with the Microsoft paperclip help agent is similarly disassociated from the animated presentation (as shown in Figure 6). A similar analysis for this agent can be undertaken showing that the animation does not establish a mental model of the abilities or limitations of the dialogue or knowledge of the agent - indeed it also appears to change expression, or move its eyes around for no reason in the same way that the Autonomy hound "shuffles".

A further criticism is that when the states are learned by the user, they are mostly irrelevant to the user's task (Requirements 4, 5 & 6). The user does not need to know whether the assistant is "running to a web site" or "reading a web page" beyond the reassurance that activity is taking place, since these are not activities in the user's model of the task which affect interaction. Consequently, all this index is doing is serving the function of an elaborate active process index which supports the marketing image of the product. Indeed because the index is actively "shuffling" even when the process is inactive (Requirement 7), it does not even serve as an active process index cleanly.

Figure 7: The Autonomy WWW agent uses pop-up dialogue boxes for training

This analysis of an example animated interface agent does not suggest that it meets the requirements well. Indeed the introduction of a character does not seem to support a metaphor that matches the user's mental model of the task in order to improve interaction, beyond limiting the responsibility which is likely to be delegated to it.

It is clearly attractive to introduce the notion of character in order to overcome this major problem, but what would make a good character, and representation ? In Autonomy it can be argued that the metaphor of the animated hound is not a metaphor to real dogs in the real world, but an analogy to the anthropomorphised animals we experience in animated cartoons on television and in the cinema. If this is so then the skills of film animators should be used to define what is a good characterisation or presentation. From film and animation craft we can derive the following guiding principles for animation (from Young and Clanton, 1993):

Appeal
The image should be pleasing to the eye
Simplicity, communication and magnetism
Avoid poor legibility, over complex, awkward or poorly designed images
Avoid misinterpretably delicate expressions
Avoid extreme close-ups that flatten objects by eliminating details that give substance
Exaggeration
Find the spine in the story, its essence, and exaggerate
Caricature reality to be more convincing
Exaggerate too much, back of later if required
Especially during repeated actions, extreme exaggeration clarifies action and is unnoticeable
Staging
Clearly and effectively present important ideas to make the story point that advances the plot
Plan the presentation of each significant idea for maximum clarity and impact
Consider variations of camera angle, type of shot, cuts, production values etc..
Eliminate confounding actions, unnecessary background
Focus attention with close-ups, lighting etc.
Timing
The timing of an action creates its expression
Time is very natural in interaction
Synchronization of sound and visual action must be frame accurate
Realism
Objects should have weight, depth, balance
Objects should be plastic- strong not rigid
Feathers should be light, a lamp base heavy
Realism does not imply naturalism.
The animated world should be believable but not true to life.
Do not alter the physics of the world arbitrarily, only for special effects.

Although these craft principles are useful within the context of film animation, film craft does not contain an overall theory that can guide us in their application outside the film itself. Again it is tempting to turn to film and animation semiotics for a theory to guide to the design of animated film characters since these areas have been investigating animation for many years. Metz (1974) and following works have presented semiotic analyses of film and animation which address narrative pace and time issues in great detail which are beginning to be further developed in terms of human computer interaction (e.g. May and Barnard, 1995). On issues of character the theory is less helpful since structuralism generally permits the reading of symbols in many ways, so much so, that the many interpretations of figurative readings is the main content of structuralist criticism. Metaphors are groundless in that they permit the substitution of one set of signs for another, and language betrays its own fictive nature just at those points when it is offering to be most intensively persuasive. Paul de Man and the post-structuralists go so far as to state that literature acknowledges its own rhetorical stature in that all claims to knowledge are made without a theory of figurative structures which render them ambiguous and indeterminate. Other forms of discourse do not always make this acknowledgment. In the case of film, Metz argued that film is a language rather than a language system (langue) since it does not support two way communication between the film and the viewer, only deferred communication one way from the film maker, through the presentation to the viewer. This argument has led to counter-arguments that it is centered on speech communication rather than general language use; but even post-structuralist textual critics have tried to describe textual reading as a dialogue between the author and the reader through the text. They argue that the author proposes ideas to the reader, and the readers reactions or questions to these must be answered by reading the text further. This limited view of dialogue in film and literature provides a limitation on the influence we should take from these media into dialogue design for human computer interaction.

Film is not a language system, since unlike natural language systems it is only one way. Similarly the rules for character formation in film or advertising are limited to this one way communication - the viewer or reader does not have to develop a model of the characters plans and language comprehension and map it to their own as they do with another discussant in a dialogue. With interactive characters this is not so. Such modeling and mapping is required as it is with the HCI natural language interaction style, where NL interaction mechanisms for disambiguation and clarification are required when the possible interpretations of the signs become manyfold because of the complexity of their interactions. Is it possible to take the principles of the craft skill of animation to produce an exaggerated character who is still sufficiently minimalist so as to avoid these complexities ? This is what has been attempted with the Autonomy hound, but it fails to work since the character tells us nothing about the communicative abilities, or the knowledge embodied in the underlying assistant, no matter how successful the character is at limiting our expectations.

6 Characterisation and Realism

The Autonomy agent is assumed to be a analogous to a cartoon dog rather than a real dog. It has features exaggerated to emphasise its unreality, and promote the cartoon rather than real world analog using the principles of animation above. Since the animated hound is not one which supports a language system and is not one for which viewers normally develop models of plans or language comprehension ability. This is clearly an attempt to produce a metaphor to something which can includes intention, but is also unreal. Is this the right metaphor to be using, or should a more realistic human analog be found ?

Source	% Information
Physical Appearance	20
Body Movement & Posture	15
Facial Expression	15
Eye contact	30
Intonation	10
Word Choice & Grammar	5
Propositional Content of Speech	5

Table 1: The percentage of information from different sources understood about a person on first meeting them (recalled from memory, so the numbers are inaccurate).

Table 1 shows a classic social psychology result describing the source of information understood by a person on first meeting another. If a more realistic analog were used then the analytic comprehension processes reflected in this result would be brought to bear on the animated characterisation. To produce such an analog requires considerable complexity since these factors interact in complex and very time dependent ways which are very intolerant of asynchrony. Some examples of the roles of these non-propositional cues are listed below from Pelachaud et al (1996) :

Body movements
   At change of conversational topic
   At turn taken, after eye contact changes
Head movements
maintaining flow of conversation as a turn taking cue
nodding for agreement
shaking for disagreement
head direction downward may indicate sadness
head direction may point at something
speed of head movement is affect-dependent
Postural Shifts (wide amplitude linear shifts) at turn taking
Postural Shifts at grammatical pauses
Facial Expression
   synchronized with phoneme level
   eyebrow movement synchronized with word boundary
   facial movement synchronized with phoneme boundary
   happiness affect - smile, raised cheeks, wrinkles around eyes
   disgust affect - nose wrinkling, raised upper lip
   sadness affect - wrinkled brow, down turned mouth & eyes
   angry affect - frowning brow,
   fear affect - abnormally opened eyes
   surprise affect - raised brow
   amplitude and frequency of affective movements varies with arousal
Eye Behaviour
   Eyes scan objects of interest with longer glances
   Eyes scan discussant's face - 58% eyes, 13% mouth, 1% other face
Eye Contact
   Avoiding eye contact is variously interpretable
   eye contact aids conversational synchronization
   eye contact is synchronized with intonation contours
   At end of turn eye contact is temporarily broken then re-established
Eye Blinks
   Eyes blink every 4.8 sec, & last 0.25 sec, with 0.04 sec closure
   Blinks are used to accentuate words & mark pauses
Intonation
   Synchronised with body movement, gesture & facial expression
   At the end of turn intonation drops and rises - normally

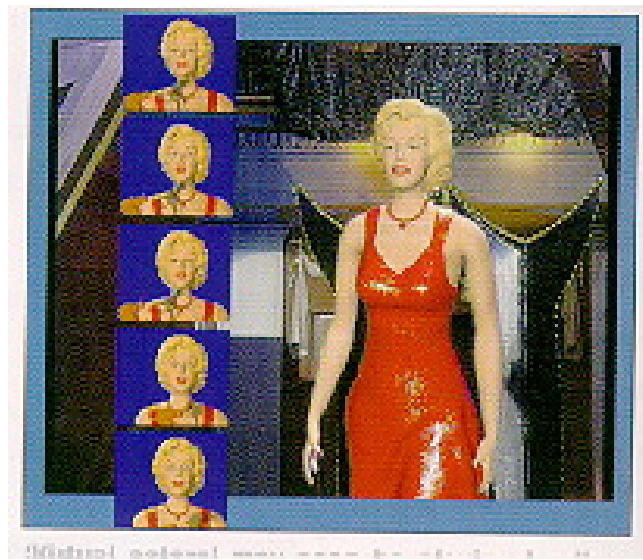

Figure 8: The HUMANOID II animated figure in a virtual world.

Although there are developments (e.g. Pennsylvania University's Jack - illustrated in Figure 9) of interactive humanlike presentations which could be used as the animated presentations of intelligent agents which attempt to incorporate these features and the supporting synchronisation required, they are not yet close to mimicking the human presentation system and result in sufficient mismatches to make their use problematic. Several studies have shown that although the use of human faces appear to increase engagement, they also take more effort to interact with the system (Koda and Maes, 1996), lowers performance and lead to the user generating too high expectations of the systems abilities (Takeuchi et al, 1995; Walker et al, 1994).

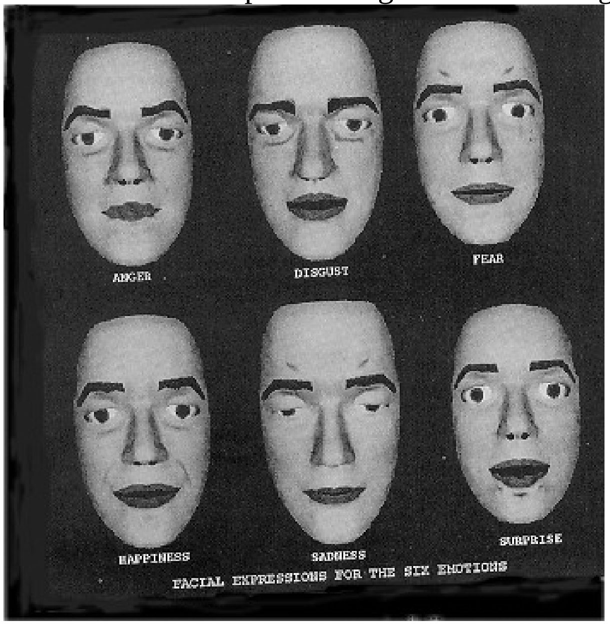

Figure 9: Animated Faces from Jack, expressing affects, synchronised with speech.

The biggest problem with developing human-like representations, as with the generation of progressively more complex representations of any class is that users assume that parts of the display have been presented to convey some information which in fact was not intended (Reiter & Marks, 1990) because there is information in the presentation which is redundant to the message, indeed worse than this, the extra information reduces the quality of the presented information. It is important to maintain control of the complexity of the presentation within that which the generation system can actually plan, and either be certain of how it will be understood, or provide dialogue mechanisms for clarification and disambiguation. Applying this constraint, the first interactive game from Dreamworks which applies the craft skills of Hollywood animation to computer games uses minimally expressive clay figures for characters built to avoid false implicatures by only introducing elements whose potential role as indicators of character psychology have been planned and controlled - illustrated in Figure 10.

The animated agents discussed so far have been limited to desktop agents which open to provide a different interaction style to support communication. Animated presentations such as PEEDY from Microsoft Research (Ball et al, 1996) is an animated desktop agent that uses speech production and comprehension based on a cartoon parrot metaphor. In this case the parrot is probably the cartoon character most associated with simple speech ability, and this choice again supports a tolerance for failure in the communication skills while also indicating that speech is most likely as the communication medium.

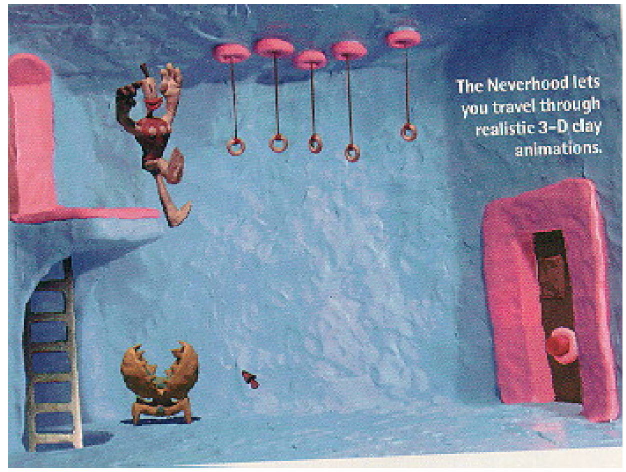

Figure 10: Clay men in the first interactive adventure game from Dreamworks - Neverhood.

The next progression of animation is to move from 2D to 3D VR where the agent can interact with representations of users (Benford et al, 1995; Bowers et al, 1996) as well as other agents and instruments. There are currently many developments in the technology for animating such agents (Zhukov and Iones, 1997; Zeltzer and Johnson, 1994; Thalmann, 1994, illustrated in figure 8) and standards in preparation for the virtual environments in which they will operate (e.g. The living Worlds proposal for VRML 3.0). In such an animated environment it would be possible for agent representations to interact with each other and express their nature as Collaborative or Selfish Agents (e.g. Brainov, 1996). However, the ten simple requirements derived from the simple examples at the start of this paper will still apply.

7 Conclusion

Ten simple requirements have been derived which animated representations of agents at the user interface should meet :

1) Delegate tasks and responsibilities to the agent - the interface must afford user actions to do this.
2) Evaluate the execution of a task in order to decide to terminate the delegation.
3) Explain the results of its reasoning., and from that explanation, decide on the competence of the agent sufficiently to assess whether to delegate further tasks.
4) To promote a mental model in the user consistent with the abilities of the assistant and compatible with the user's representation of the task.
5) icons to convey existence and metaphoric semantics of state
6) indexes to convey events and state change
7) indexes to convey indivisible process activity
indexes to convey indivisible process duration
symbols to disambiguate and clarify the other signs
affordance of user actions available.

The representation of the agent does have to support the marketing of the tool as well as the interaction, but this need should not swamp these simple requirements, otherwise initial sales may be high, but users will not be able to use the tool, and the product will not continue to sell.

In order to overcome the "Guru" role of expert systems it is necessary to develop a mental model in the minds of users which clearly limits the expected intelligence and communication skills of the agent. A simple cartoon representation appears to be able to do this, when it is designed to be analogous to cartoon characters rather than real world ones. The closer it gets to a real character, then the more requirements for communication skills are placed on it, and the less scope for the exaggeration and breaking of the rules of the natural world will be permitted or supported in the user's mental model.

References

Autonomy Corporation (1996) http://www.agentware.com Cambridge, UK

Ball et al (1996) Lifelike Computer Characters: the Persona project at Microsoft Research. In J. Bradshaw (Ed.) Software Agents, MIT Press.

Benford, S., Bowers, J., Fahlen, L.E., Greenhalgh, C., Snowdon, D. (1995) User Embodiment in Collaborative Virtual Environments, Proceedings of CHI '95, 242-249, ACM Press: New York.

Bowers, J., Pycock, J., O'Brien, J. (1996) Talk and Embodiment in Collaborative Virtual Environments, in Proceedings of CHI '96, 58-65, ACM Press: New York.

Boy, G. (1991) Intelligent Assistant Systems, Academic Press: London.

Brainov, S (1996) Altruistic Cooperation Between Self-Interested Agents, in W. Whalster (Ed.) Proc. ECAI '96, 519-523, Wiley: Chichester.

Koda, T and Maes, P, (1996) Agents and Faces : The Effects of personification of Agents. Poster at HCI '96, the annual conference of the BCS HCI group, August 1996.

Living Worlds (1997) Living Worlds: A Framework for the Creation of Multi-User Virtual Worlds., http://www.livingworlds.com/draft_1/

Maes, P. ((1994) Agents that Reduce Work and Information Overload, Communications of the ACM, 37(7), 31-40.

May, J. and Barnard, P. (1995) Cinematography and Interface Design, in Proceedings of IFIP Interact '95, Lillehammer, June 1995.

Metz, C (1974) Language and Cinema, The Hague: Mouton.

Newman and Lamming (1995) Interactive System Design, Addison-Wesley. Wokingham.

Pelachaud, C., Badler, N.I., and Steedman, M. (1996) Generating Facial Expressions for Speech, Cognitive Science, 20(1) 1-46.

Reiter, E. And Marks, J. (1990) Avoiding Unwanted Conversational Implicatures in Text and Graphics. In Proceedings of AAAI '90, Boston, July 1990.

Stam, R., Burgoyne, R., Flitterman-Lewis, S. (1992) New Vocabularies in Film Semiotics, Routledge: London.

Takeuchi, A. Et al (1995) Situated Facial Displays: Towards Social Interactions, In Proceedings of CHI '95, 450-454.

Thalmann, D. (1994) Automatic Control and Behaviour of Virtual Actors, in L. MacDonald and J. Vince (Eds) Interacting with Virtual Environments, Wiley:Chichester.

Walker et al, (1994) Using a Human Face in an interface, In Proceedings of CHI'94, 85-91, ACM Press: New York.

Young, E. And Clanton, C. (1993) Film Craft in User Interface Design. Tutorial at IFIP & ACM InterCHI'93, Amsterdam.

Zeltzer, D and Johnson, M.B. (1994) Virtual Actors and Virtual Environments, in L. MacDonald and J. Vince (Eds) Interacting with Virtual Environments, Wiley:Chichester.

Zhukov, S. And Iones, A. (1997) Real-time Behavioural Control and Procedural Animation of Intelligent 3D Agents, in Proc. 15th Annual Conference of the UK Chapter of Eurographics, Norwich, March 1997