Designing a Non-Verbal Language for Expressive Avatars

Ben Salem*
University of Plymouth
Exeter School of Art & Design
Earl Richard Road
Exeter EX2 6AS England
+44 1392 475126
electronote@mail.com

Nic Earle
University of Plymouth
Exeter School of Art & Design
Earl Richard Road
Exeter EX2 6AS England
+44 1392 475126
nic_earle@postmaster.co.uk

ABSTRACT
Collaborative Virtual Environments (CVEs) were designed as an expansion of the text-based chat room, rather than a novel application, exploiting the possibilities of online three dimensional graphical space. This initial design direction is observable at the interface level. We put forward the case that to achieve an efficient CVE system, one will have to design and implement a multi modal User Interface based on expressive Avatars as a representation of the different participants, also as an embodiment of software agents. We emphasise the expressiveness of the avatar as a crucial improvement to the efficiency of their communication capabilities, and we describe a vocabulary of expressions to be implemented. We put forward the case that to be more efficient, particularly during a dialogue, an avatar is required to play a role in the communication using non-verbal channels such as body postures, facial expressions and hand gestures. We also suggest conversation circles to facilitate the gathering of participants in a discussion. These circles will address navigation difficulties in CVEs and encourage social exchanges.

Keywords
Avatars, Facial Expressions, Hand Gestures, CVE, Social Communication

INTRODUCTION
A Collaborative Virtual Environment (CVE) is a multi-user virtual space where users are represented by a 3D image termed ‘avatar’. The participants of a CVE should be directly visible to themselves and to other participants [1]. This creates a more immersive experience and has greater potential to provide important information to a user about their avatars surroundings as well as reinforcing the perception of the environment. Several papers have been published on the visual aspects of avatars however few have addressed theirs communicative functions [1][2][3]. The issues we wish to address are not only related to the characteristics of the avatars in terms of shape and visual appearance, but are about their functions and behaviours in relation to their owner.

Within this paper the term 'avatar' is also used to describe a graphical representation of a software agent which some CVEs feature. Actions of this type of avatar are controlled by software. The advantage of giving an agent a embodiment is to firstly humanise the agent and secondly to allow a more natural method of interacting with it.

The use of common conventions, that one is familiar with, facilitates the interaction within the CVE as well as improves the communication between users, and between users and agents. Humanoid avatars for the participants are therefore the most efficient design option, they allow the use of everyday conventions such as gestures, postures and body language as part of the communication channels between users. For the embodiment of agents there is more freedom about the realism, and type of avatars to be used while keeping a practical level of expressiveness and interactivity. The use of abstract or symbolic representations for a participant is less justifiable if one wants to achieve a user-friendly environment delivering a rich experience with a degree of multi-modal communication. Having a human-like appearance makes it easy to gather the aim of a conversation, as well as to recognise the participant to whom one is speaking. There is however a need to draw on these principles and develop a set of design guidelines that are useful to the development of avatars and CVEs.

A further problem in developing a non-verbal language of avatars is that avatars that are in conversation are usually scattered around an environment. Even avatars that are within the same conversation are often out of view of one another, forcing the participants to rely on the text window. This is a major problem as any gestures and expression that are performed are likely to be missed. There is a need for the design of a better conversation space which we suggest in this paper.

THE ROLE OF AVATARS

The role of avatars in CVEs is essentially to represent a user's presence, orientation and location. We propose a set of design guidelines to deliver avatars that have further capabilities, namely to facilitate the communication of ideas and the expression of emotions as well as identifying its owner. We address issues concerning group communication, personal identity and expressions. There is a need for avatars to convey relevant information about their users real identity and his intentions during discourse and conversation. Our research problem therefore divides into two main areas. The first aims to raise issues that need to be addressed about personal identity of human participants. The second is concerned with aspects of the individual and group communication for both human to human collaboration as well as human to agent collaboration.

In the latest group there has been a steady increase in the use of software agents for the assistance and guidance of new participants in a CVE unfortunately this has been done without a common set of behavioural rules and vocabulary of expressions. Leading, to a difficulty in understanding the agents, or to the false impression that one is facing another user.

Figure 1 : Anthropomorphic avatars (from Active Worlds)

Some of the avatars used in CVEs are humanoid but inadequately designed, as shown in Figure 1, the level of detail is either too poor to permit rich facial expressions, or the design is such that few expressions and gestures are possible.

Facilitating the User's Discourse and Interaction

Once logged-in, communication between the CVE participants is possible by text chat and a library of body and hand gestures. Text messages appear on the screen either above the avatar of the message author or within a separate dialogue window. Users are able to interact with the agent software through the exchange of text messages.

Figure 2 : High level of detail Avatar (from Poser).

Avatars can be realistic, abstract or naturalistic in form. We ask what is the best style for CVEs. Realistic avatars strive to provide an accurate representation of the user. For example a participant with a realistic avatar may use their real name as a username and real time video image of their face wrapped onto their avatar face [2] [4].

Naturalistic avatars are usually humanoid in form, of a degraded level of details, but emulate natural paradigms just enough to achieve recognition of familiar features [5]. We advocate the use of this style in CVEs. As a humanoid form it will open up the possibility of using conventional conversational habits, such as looking at the speaker, and using all the non-verbal communication techniques one has learnt to employ during a conversation.

Much research has gone into realistic modelling of virtual humans. Notable success are the JACK system [6] and the Xian Terracotta Soldiers [7], however this is a difficult path paved with problems when attempting to apply results to 3D avatars. This is due to technical restrictions such as bandwidth and computational power.

Despite several years as a research topic, avatars in existing collaborative systems play little role in the communication that takes place. Communication is achieved almost entirely on the exchange of text or in some cases voice messages, while in the real world communication is far richer then this. It involves other non verbal channels of communication (NVC) such as face expression, eyes' gaze, eye opening, duration of glances, facial expressions. As well as body postures and gestures, which makes the difference between formal and casual conversations. Also to be included are hand postures and gestures as part of the discussion, for example, to emphasise words, and quantify them, body contacts, hand shakings and slap on the back.... Finally, Clothes, makeup and other aspects of appearance, formal and casual dresses and, uniforms play a significant role as well [8][2][9]. NVC play an important part in our social interaction. When several channels of communication work together we use the term 'multi-modal'. With a human-like avatar,

it is possible to exploit some of the non-verbal communication (NVCs) capabilities, facial expressions, hand gestures and, body postures. At the same time these NVCs will be perceived and easily understood by the other participants in the CVE.

A relevant system known as Comic Chat developed by Microsoft integrated the visual language of comic strips with text chat [16]. The result was a brilliant system where symbols were extracted from text messages to automatically generate body gestures, face expression and background scenery. While some useful lessons can be learned from the project its carton like aesthetic encourages anonymity and playful behaviour. It is therefore more suited to chat and role playing games than CVE applications.

Existing CVEs

In existing systems however there are usually two separate modes of operation. The first is controlling an avatar to get into a spatial position where one is able to enter discussion. The other is typing text or speaking, which in current systems implies a halt to the control of the avatar. In order to develop a multi-modal communication, a visual language for the avatar needs to be developed that supports the discourse that takes place, with a merging between the two. This is particularly true for CVEs that that have a separate window for the textual dialogue. Such a separation removes the user's attention from the avatars and the environment, in which case any visual clues performed by an avatar would be missed.

A final problem concerning social communication that is addressed within this research is that of activation. Some existing virtual worlds such as the commercial system Active worlds offer a limited number of gestures such as waving and some body expressions such as Laughing. However these gesture need to be activated by selecting buttons from tools bars within the virtual world software. We believe this to be totally inadequate solution. In the real world we use gesture in an effortless manner. Often we perform them subconsciously and even involuntary. This makes explicit activation of them from menus completely inadequate.

AVATARS KEY ROLES

We suggest that avatars are designed to address keys issues in the interaction process between the user and the CVE. We classify the key issues as the identity of the owner, the accessibility of the owner, the status of the owner, and the avatar functions and characteristics.

Identity of the User

Within the limits specified by the owner, it should be possible to gather as much information as one would wish about an avatar's owner such as identity, gender, location, interests.

Accessibility of the Owner

Has the owner got a web cam, an audio link, or is text chat the only possible link?

Status of the Owner

Is the owner of the avatar a guest user, a privileged user, a manager who is monitoring the CVE. Is the user on pause mode, how long has he been silent, on a poor internet connection. Does an avatar represent an agent or a human (is this important?).

Avatars Functions and Characteristics

What are the actions within the CVE that can be performed by the avatar, dialogues, displacements, manipulations, expressions, gestures.

Some multi-user environments such as the commercial Blaxxun world [10] allow participants to use more symbolic avatars. These avatars are abstract and usually encourage the real participant to remain completely anonymous. Participants with a symbolic avatars have the tendency not to use their real names when assigning themselves a username. Symbolic avatars encourage anonymity which tend to eliminates shyness and facilitate purely casual chats. This is not appropriate for most applications of CVEs. Furthermore the development of a gestural language for this type of avatar can not be based on the assumption that the gestures will be understood. Due to the abstract and symbolic nature of the avatar any suggested gesture or expression may be far too remote from the familiar ones to be easily perceived let alone understood.

GESTURES & EXPRESSIONS FOR AVATARS

There are several ways of performing gestures and expressions for avatars, which include the following: Real time manipulation of avatar limbs and face using the participant real body connected to sensors. Real time streamed video of a participant onto their avatar. Calling up predefined animations from a local library of gestures and expressions.

The first two methods are computational expensive and require specialist equipment. They also have a major problem in that the participant body resides in a very different space to their virtual body. As a example if a participant wishes to turn around to see who is standing behind them, they will have to turn away from the screen. One of the drawbacks of video streaming is the impossibility of achieving eye contact. To look someone in the eyes, requires looking at the camera instead of the screen and consequently not paying attention to the environment anymore. These configurations also do not transfer well to desktop systems. We therefore favour the predefined library of gestures and expressions.

Within this approach there are several issues that need to be addressed which are as follows:

1. What is the best form of NVC i.e. gesture or facial and body expression.
2. How do we arrive at a set of useful gestures or expression for both human participants as well as agents?
3. How can we ensure that the gestures and expressions are not missed by others when performed? Or how do we ensure that the communication becomes truly multi-modal.
4. How can they be triggered in a more natural way than selection from a tool bar?
5. How can we ensure that they are synchronised with the correct word or message?

What is the best form of NVC

The first issue to address is the choice between using facial expressions or body and hand gestures. In the real world they play slightly different roles. The face shows emotion and punctuates words and sounds while the body shows mood and status. In the virtual world however these two uses do not easily translate. While users engage in one to one conversation with another person, their head and shoulders may fill the screen. This results in facial expression being noticeable while body and hand gesture are out of view. This is reversed when two people are communicating with a relatively large distance between them or when participants are conversing within a group of people who's avatars are gathered together. In this situation facial expression can be too subtle and are easily missed. Many styles of stage performance in history such as Commedia Dell Arte [11] give the actors masks to hide expression and encourage the use of expressive behaviour via the body. This principle is useful in this type of interaction.

We therefore propose that there should be two sets of animations developed for each meaning that is to be visually conveyed. One set is implemented through gesture, the other is through facial expression. These sets load up when required, based on proximity. When two avatars are within a certain range, the facial expression set loads. When they move out of range the gestural set loads.

This raises the question what will others in the environment see when they look over at a group from a distance? This should depend on their actual distance. It should be gestures unless they are close enough to view the face.

Body Postures

Body postures can be used to express the general state of the avatar. For example during a phone call the participant may want his/her avatar to express a temporary idle state.

Facial Expressions

Facial expressions are an efficient carrier of emotions. They are generally used to punctuate discourse at an informal personal level. When used in a context of a CVE application, facial expressions are less of an important feature for a participant avatar, however they keep all their usefulness for agent avatars independently of the application. A powerful channel of NVC is gaze. This is different to facial expression although it involves facial features. We have found from experience that a crude simulation of gaze can be achieve through the heads direction of face.

The body direction of face can be used to show group membership, and the head can be used to simulate gaze within the group.

Figure 3 : Blinking face (from Poser)

Hand Gestures

Hand gestures are the most comprehensive NVC skill. We can classify gestures as being either :

Symbolic:

These are context independent expressions such as signs (e.g. sign language), the message is contained fully into the gesture performed.

Metaphorical:

Used in parallel to a spoken or typed message for quantification, expression of virtual and interface objects (e.g. a handle and geographical navigation), the message, in this case, is gathered from the gesture itself and the context within which it is performed. An example is using the hand to specify the intensity of the light, by moving it up and down.

Pointing:

To select objects or to reduce the scope of a spoken statement (e.g. this, that), this is a special gesture, since it can have so many meanings, most of them context dependants. Pointing could be performed to indicate a direction of displacement, an object or an option selection.

GENERATING THE EXPRESSIONS, POSTURES AND GESTURES

The generation of the facial expressions, body postures and hand gestures, is driven by key parameters, the current context within the CVE, the interaction task being performed, the current discourse of the participant if applicable.

We are primarily interested in two forms of NVC. These are hand gestures and facial expression. This is because they possess the widest range of capabilities for communication. We emphasise hand gestures since our principal way of interaction with the world where we live is through our hands; we perform almost all of everyday tasks with them, with a great dexterity and naturalness [12].

Two sorts of expressions are required: to illustrate the avatar status and to punctuate an avatar discourse. The first group is essentially made up of body postures (i.e. idle, thinking), while the second group is made of facial expressions and hand gestures. They are generally performed in parallel to a spoken or text-based dialogue.

Within textual message users often use text symbols to convey emotion and describe a bodily action. This is an attempt to replace the body in interaction. It is therefore logical that these textual symbols are given a visual clue as to their meaning. These symbols include acronyms, emotion icons and performative words.

- *Acronyms* – are abbreviation that make up a common expression e.g. LOL = laughing out loudly
- *Emotion Icons* – keyboard symbols used to create faces e.g. :-) = smile
- *Performative words* – these are words embedded within text message to describe actions. These words are usually identified by being contained within two symbols such as the asterisk e.g. *wave*

Text messages contain as well elements from traditional grammar that are used to punctuate messages. For example the question mark indicates a request for feedback and the exclamation mark emphasises a message, capitals letters are used for shouting. All these need to have a visual interpretation. There are also a number of words that are so commonly accompanied with a gesture in real world conversation that we believe they should be included within the set such as the words 'yes' and 'no'.

Examples from this section are shown in table 1:

Text	Meaning	Gesture
Yes	Agree / Yes	Nod head
No	Disagree / No	Shake head
?	Questioning	Head back, one eyebrow raised, hand out stretched
!	Emphasise message	Head back, eyebrows raised, torso upright
:-) LOL	Happy / Smiling Laugh Out Loudly	Smile Laughing
:-(	Sad / Upset	Head and shoulders dropped.
AFK	Away From Keyboard	Busy
OMG	Oh My God	Surprise
IMHO	In My Humble Opinion	Neutral pose
:-* *Kiss*	Kiss	Blow a kiss
wave CUL8R	Hello/goodbye See You Later	wave

Table 1: Examples of keywords with their visual clue

Figure 4: Waiving and Thinking/busy poses (from Poser)

TRIGGERING GESTURES & EXPRESSIONS

We propose that a more natural method to activate gestures is through the text messages that are written and sent. A system needs to be developed where the text messages are scanned for keywords. When detected these keywords trigger an appropriate gesture. A similar approach could be adopted for spoken discourse. A speech recognition module, would extract key words from the discourse. This kind of system has already been implemented in the 3D Online Traveller world where verbal messages are scanned for phonemes that activate lip synchronisation [13].

Acronyms used within CVE discourse can be directly mapped onto their corresponding gesture. This is also true of punctuation marks used in traditional grammar as well as emotion icons.

Another technique that can be used is when a user sends a message intended for a particular group member by including their nickname within a message. The software should detect this and turn the author's avatar head towards the intended recipient's avatar.

A user should also be able to inform the system, whether a trigger word or symbol is displayed within the message on the other group members screens. This is so that a gesture can be performed without a text message. For example a user may wish to indicate a desire to speak without

explicitly interrupting the person who has the floor. This can be achieved by typing a symbol such as a forward slash (/) as a message. The gesture will be triggered but there will be no text sent.

Another useful feature is to allow the user to customise what gesture is assigned to what keyword. Changes can be saved as a separate set which can be loaded up at any time. This is useful for developing different pallets or gestures that are called up for different contexts i.e. social or formal. Some gestures such as signalling an intent to leave a group conversation need to have there own key assigned. For example the function keys can be used to activate appropriate gestures.

A MULTI-MODAL CONVERSATION SPACE

Investigations into the structure of virtual world social encounters reveals that the process of interaction typically breaks down into three sequential stages [9][14]. In the start part people seek an individual or a group for meaningful conversation. In the middle users interact and in the end they negotiate a way breaking out of the interaction.

In current CVEs avatars are scattered all around the environment. This distribution of the avatars is a direct consequence of difficulty in navigating in the CVEs. Another reason for the lack of clustering, one would expect when a conversation occurs, is due to the fact that although each participant's text message appear above their avatar in a bubble text, the text is often difficult to make out. This is even more the case in a crowded environment where text messages tend to overlap each other. Participants have the tendency of relying on the text window, and the experience shift from participating into a CVE to taking part in a text chat session. This itself has further reduced the need for participants to move their avatar into a gathering when engaged in a group conversation.

In order to achieve multi-modal conversation, not only should other channels of communication be developed such as gestures, but users avatars must be in a space where they can see the other group members. We believe that the need for two windows i.e. text and visual, can be eliminated. Text message can appear in the same space as visual ones. This is a necessary step toward achieving a multi-modal interaction.

We have developed a concept we call the ‘conversation circle’. This is a configuration for achieving multi-modal communication with chat systems. Users have been reported to arrange their avatar into a circle formation for discussion in existing virtual worlds [15]. Our proposal builds on this formation and presents a more sophisticated application. Figure 5 shows a screen shot of six avatars within a conversation circle. A circle begins when one user makes an offer of conversation by clicking on the recipient avatar. This is detected by the computer and the users avatar is animated over to the recipient on all user monitors and an appropriate gesture is triggered. If the recipient accepts the offer of conversation by clicking on the users avatar, both avatars are placed and locked along the circumference of an invisible ring. A user must signal that the conversation is over before they can leave the ring and move on. This can be achieved by detecting that a mutual obligation has being meet to leave the conversation for example, the wave gesture or a textual symbol embedded with the final text message sent to the group such as a forward slash (/).

When a group is formed an invisible light fades in from the circle centre and just above head height. This light illuminates the avatars faces. It also creates shadows around the parameter which helps to visually define the groups space. Users external to the group can not move in this area. When last two members of a group end their interaction the light source fades out.

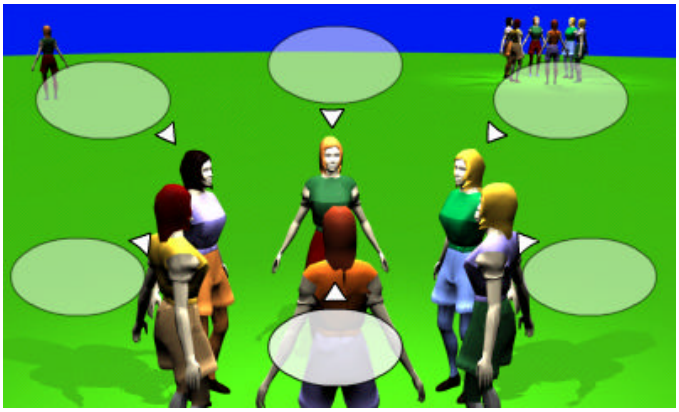

Figure 5: Screen Shot of a conversation Circle

Newcomers can join the group by clicking on it with the mouse. The circle expands and the newcomer is placed along its circumference. Avatars are evenly distributed along circumference of the circle so in terms of spatial importance all users are equal. When a user departs, the circle contracts to bring together the remaining avatars.

The users viewpoint over the group needs to be third person view so that all participants are visible. This includes a users own avatar which is important as s/he must be aware of what gestures their own avatar is performing.

Text messages appear in speech bubbles that emanate from the avatars. They remain on screen for a limited duration determined by the length of the message in characters, or until a new message is submitted by that user. Text message within other groups are not visible while in a conversation circle as other discussions are not considered

to be relevant. This also ensures that the messages are not overlapped. The text bubbles themselves are opaque so as not to completely obscure the environment.

This concept is still in its early stages. Although multiple social circles can exists within a single environment, each circle currently only accommodates for up to six group members. This particular formation has been designed with purely chat applications in mind however we believe that it can be modified to support other uses of CVEs. A final point to note is that few CVEs currently support the use of shadows. It is however possible to simulate them with the use of algorithms. This technique is commonly used within computer games such as Tomb Raider which is available for the Sony Playstation and the PC.

The Convention of the Conversation Circles

Stage 1: Initiation or joining a conversation

Table 2 shows some gesture that can help visually support or this part of the interaction.

Meaning to Convey	**Visual Gesture**
Available for conversation	Remain Neutral
Unavailable for conversation	Ghosted avatar (semi-transparent effect)
Ask for conversation (close/far)	Move into a space, meet gaze, perform questioning gesture
Accept an offer or conversation (close or far)	Smile/move towards the conversation space

Table 2: Table of Gestures to support the initiating of a social interaction.

Stage 2: Engaged within Conversation

Once engaged in conversation there are several visual operators that can help facilitate fluent conversation. examples of these are shown in table 3

Operator	**Visual Gesture**
Greeting	Wave a hand
Give floor	Arms forward offering, then pulled back /offer token
Show who has floor	Hold token
Show desire to speak	Raise arm above head

Table 3: Gestures to ensure the smooth running of conversation at the middle part of an interaction.

Stage 3: Leaving Conversation

Users usually signal an intent to end an interaction some time before it actually ends. The interaction will then end only when both participants or in the case of a group, all participants, have come to a mutual agreement.

Signal	**Gesture**
Signal intent to leave	Body quarter turn, gaze still connected.
Say goodbye	wave

Table 4: Necessary visual clues for ending conversation

Some gestures need to be adapted and there is a need for some additional gestures for when a user joins and then participates within a group interaction. Examples are shown is table 5.

Joining	
Ask for conversation	Move onto circle parameter
In Conversation	
Show newcomer so that others can involve him	Avatar is highlighted
Show that a message is addressed to the whole group	Sweeping gaze and arm with hand out stretched.
Show a message addresses the an individual group member	Head turn towards
Leaving	
Signal to group mutual agreement has been recognised	Sweeping wave.

Table 5: Additional Gestures Needed When Conversing Within A Group.

Agent Avatars in Conversation

The expressions and gestures that need to be performed by an agent are slightly different for the middle stage of the interaction. Within this stage gesture performed by agents should serve to inform the human participant of its status. Table two provides some examples of states that the agent needs to convey to achieve a more fluent conversation using the visual channel.

Status	**Visual Clue**
Available	Normal
Unavailable	Ghosted
Type of agent i.e. help or policing agent	Dress code
Processing a query	Thinking gesture
Cannot interpret request	Confused expression
Require more information	Questioning expression

Table 6: Examples of necessary gesture for agent avatar in conversation

For the leaving stages a simple 'goodbye' should suffice. As conversation with agents is likely to be less involved and more functional, following the conventions of social interaction in agent conversation will be less of a requirement. All that is required to break off an interaction with an agent is a single signal to show the agent that they are no longer needed for the time being.

The Participants Field Of View

When designing the conversation space one of the key issues to be addressed is the point of view from which the participants will see the circle as well as the other avatars. Three options are available a first person view, a third person view and a bird's flight view. The first person point of view is the one delivering a realistic visual observation of the environment. Such a point of view doesn't deliver a comprehensive perception of the environment, this is due to the lack of any feedback from the avatar actions (such as balance, touch and force perceptions). Without visual feedback, one cannot be fully aware of what the avatar is currently doing with a first person view [3]. A third person view is basically located behind the avatar slightly above the head, this lets the user not only see what the avatar is looking at but at the same time be aware of the avatar actions, postures and immediate surrounding. Because this view is fixed relatively to the avatar no disorientation can occur when there is a mismatch between the avatar gaze and the user's point of view. Unfortunately this is the case with the bird's flight point of view. The user has to be aware of this fact and not attempt displacements when such mismatch occurs. With the third person view it is possible to have all the participants in the circle of conversation within the field of view as one would expect it to be the case. With a bird's eyes point of view, comes the possibility of viewing more than just the local conversation circle. As a thumb of rule, first person view is to be used only for formal or urgent one to one dialogues, third person view in most cases, and bird's eye view when inquiring about the environment.

GROUP STATUS

We argue that there is a group behaviour that needs to be visually represented. The social circle concept allows for a visual grammar of a social group to be developed. For example it may be necessary to provide a visual clue that indicates whether or not the group is full. If a group is full then the light source can turn red. Other users can then understand at a glance the state of the group.

Social circles are owned by the participants contained within them. This means that only group members can change the status of the group which could be :

- Private : Invited participants only.
- Semi-private : Only acquainted participants can join.
- Public : open to all.

From another perspective conversation spaces are characterised by the type of conversation they are holding. Casual conversations are the least regulated, and no control or restriction is put upon participants regarding interruptions and entering and leaving the group. Formal conversations have more restrictive rules. Participants exchange a token which act as a moderator of the conversation.

Group Condition	**Visual Symbol**
Group is close to newcomers	Emanate a red glow
Group is open to newcomers	Normal
Group is looking for newcomers	Open spaces around the parameter
The group context is social or formal	Dress code of the avatars

Table 7: Some Visual Clues to Group Status.

CONCLUSION

We propose that to arrive at a useful set of gestures, one should draw on a number of sources. Firstly gestures that support the social conventions of a CVE, secondly from bodily actions that are described by acronyms, emotion icons and keywords that are found within the text messages.

We have Set guidelines for avatar design that take into account the identity of the owner and some of his or her relevant characteristics, and we have proposed a library of predefined gestures and expressions. Such a library should facilitate communication and expression in a CVE.

We have proposed conversation circles that address the following issues :

1. Navigating into a conversation space, allowing a clustering of conversation groups.
2. Visual definition of the conversation circles within the CVE.
3. Allows all participant to be visible at once and a multi-modal communication is achieved.
4. Delivers conversation groups which can be expanded and reduced according to the number of participants.
5. Ensures that text messages do not overlap

Further developments should result in the spatial arrangement of the conversation circles in social groups. Enhanced avatars so that members of a group are recognisable by some visual characteristics. The possibility for participants to initiate a conversation simply by clicking on an individual to initiate a group, or by clicking on an existing group which is open to newcomers.

This paper has proposed a set of design guidelines towards a truly multimodal CVE which would facilitate the communication between users and provide better ways for the participants to express themselves.

ACKNOWLEDGMENTS

This work was undergone at the now disbanded NVRCAD centre. The NVRCAD and part of this work have been funded by a JISC-JTAP grant. Poser is a software for animating 3D human figures.

* The first author is now at the Electronic Research Group, Sheffield Hallam University, City Campus, Howard Street, Sheffield, S1 1WB, England.

REFERENCES

1. Benford S., Bowers J., Fahlen L.E., Greenhalgh C., Snowdown D., Embodiments, Avatars, Clones and Agents for Multi-User, Multisensory Virtual Worlds, *Multimedia Systems*, *5*, 2, 1997, 93-104.
2. Capin, T.K., Pandzic I.S., Thalmann N., Thalmann D., Realistic Avatars and Autonomous Virtual Humans in VLNET Networked Virtual Environments, *Proceedings of From Desktop to Webtop: Virtual Environments on the Internet, WWW and networks*, (Bradford, UK, April 1997)
3. Guye-Vuillieme A., Capin T., Pandzic I., Thalmann N., Thalmann D., Nonverbal Communication Interface for Collaborative Virtual Environments, *Proceedings of CVE 98*, (Manchester, UK, June 1998), 105-112.
4. Pandzic I., Kalra P., Thalmann N., Thalmann D., Real-Time Facial Interaction, *Displays 15*,3 , 1994, 157-163.
5. Damer B., Putting a Human Face on Cyberspace: Designing Avatars and the Virtual Worlds they Live in, *Proceedings of SIGGRAPH97* (Los Angeles, CA, USA, August 1997), ACM Press, 462-464.
6. Balder N., Philips C., Webber B., Simulating Humans, Oxford University Press, Oxford UK, 1993.
7. Thalmann N., Pandizic I., Moussaly J., The Making of the Xian Terra-Cotta Solders, *Digital-Creativity*, *8*, 2; 1997; 59-66.
8. Argyl M., Bodily Communication, Second Edition, Methen & Co., Ltd., London, UK, 1975.
9. Vilhjalmsson H., Autonomous Communicative Behaviors in Avatars, available online at the following location: http://hannes.www.media.mit.edu/people/hannes/msthesis/
10. Blaxxun Virtual Worlds is a commercial product available online at the following location: http://www.blaxxun.com
11. Fisher J., The Theatre of Yesterday and Tomorrow: Commedia Dell'Arte on the Modern Stage, The Edwin Mellen Press, Lewiston.
12. Sturman D.J., Zeltzer D., A Survey of Glove-Based Input, *IEEE Computer Graphics and Applications*, Special issue on Virtual Reality, *14*, 1, 1994, 30-39.
13. Online Traveller is available at the following location: http://traveler.onlive.com/community/trav2dest.html
14. Earle N., Beardon C., The Role of Obligation Within Virtual Encounter, *Proceedings of CVE*, (Manchester, UK, June 1998) 57-66.
15. Damer, B. (1996) Inhabited Virtual Worlds, *ACM Interactions*, Sept-Oct, Vol III - 5, 27-34.
16. Kurlander, D. Skelly, T. Salesin, D Comic Chat, *Computer Graphics Proceedings, Annual Conference Series, SIGGRAPH*, (New Orleans, Louisiana, August 1996), 225-236.