A Model of Conversation Management in Virtual Rooms

Steve Benford, Adrian Bullock, Neil Cook, Paul Harvey, Rob Ingram and Ok-Ki Lee
Department of Computer Science
Communications Research Group
The University of Nottingham
University Park, Nottingham, UK

Abstract

Virtual world metaphors are becoming increasingly popular as a basis for collaborative systems. This paper describes how such metaphors might be extended to support communication on a large scale through the introduction of a *spatial model* for conversation management in virtual computer spaces. The key concepts of awareness, focus and aura are introduced and are then defined in terms of an abstract mathematical framework. The paper then outlines how this framework might be applied to a diverse range of technologies, focusing on the example of a text conferencing system in detail. Finally, the paper describes the current *CyCo* prototype at Nottingham. [1]

1. Introduction

The primary goal of this paper is to present a model for conversation management in large scale collaborative systems which are based on the increasingly popular metaphor of virtual worlds. Our model provides generic techniques for mediating communication in large groups and is intended to be applicable across a diverse range of technologies from text conferencing, through multi-media to virtual reality. The model adopts a “spatial” approach whereby people use the affordances of virtual computer spaces (called “rooms”) to control their interaction. In other words, they are able to manipulate their underlying communications system through actions such as movement, orientation, focusing and the use of various specialised tools. This is intended to contrast with existing explicit floor control policies.

In section two we briefly introduce of definition of a room as a virtual computer space in order to lay the foundation for the later modelling work. Section three then introduces the spatial model for mediating interaction. Key concepts are described and then defined more formally in terms of a mathematical framework. In section four we discuss how the model could be applied to a range of different systems. Finally, in section five we describe our current implementation, *CyCo* (Cyberspace for Cooperation), which aims to demonstrate the technical feasibility of our model.

1. This paper appeared in Proc. Telepresence’93 (Applica’93), Lille, France, April 1993.

2. Rooms as virtual spaces for interaction

We define a room as a virtual computer space within which people interact, either with each other or with the various tools which they find there. Put in a more abstract way, each room defines a particular domain of interaction. Rooms may represent task specific spaces (e.g. a room created for a specific project or meeting), more permanent organizational spaces (e.g. offices or seminar rooms) and also public spaces (e.g. foyers and corridors).

The use of virtual rooms as a metaphor for structuring computer systems is not particularly new. The Xerox Rooms system used the metaphor to structure two dimensional user interfaces [Henderson 85, Clarkson 91] and other systems utilising the rooms metaphor include VRooms [Borning 91], Cruiser [Root 88], several teleconferencing systems (e.g. [Leevers 92, Cook 91]), the Audio Windows system [Cohen 91] and multi-user recreational environments have been available for some time (e.g. MUD [Smith 92] and Lucasfilm's Habitat [Morningstar 91]. Not surprisingly, virtual rooms also feature heavily in discussions of virtual reality [Benedikt 91]. Undoubtedly the main reason that the virtual room metaphor has proved to be so popular is its strong relation to physical reality and therefore its highly intuitive nature.

Virtual rooms play several key roles in our work. They provide an ether, a continually available communication channel, allowing occupants to interact freely at all times. Rooms also provide awareness of the presence and comings and goings of all occupants and tools. This support for awareness in virtual environments is seen as a direct parallel to that already discussed for physical office environments within the domain of Computer Supported Cooperative Work (CSCW) [Lovstrand 91]. Finally, they provide access to all tools. The main advantage of locating tools in rooms is that all tools become collaborative to some degree through the ether that surrounds them. For example, I might chat to another person while standing in front of the notice board. Thus, rooms also encourage a blurring of the distinctions between asynchronous and synchronous and formal and informal communication.

3. Big rooms and the spatial metaphor

What happens as the number of occupants in a room increases beyond a small number? In a crowded room you are unlikely to be able to cope with knowing every utterance of every occupant. This implies the need for some kind of conversation management mechanism to help reduce information overload. One possible approach to such mediation is to introduce a "floor control" policy [Sarin 91]. Another approach to coordination has been evident in recent asynchronous CSCW projects which have proposed notations to model the workflow/process/activity within a group (e.g. the Coordinator [Winnograd 86], Domino [Victor 91], Chaos [Bignoli 91], Cosmos [Bowers 88] and Amigo [Pankoke 89]). Although these explicit floor control and coordination techniques may be useful in some circumstances, we believe that they are generally inappropriate for mediating the subtlety of human interaction in virtual rooms. More specifically, they suffer from trying to make explicit in the system what is essentially social convention and in so doing tend to be overly constraining. Our aim has therefore been to identify alternative mediation mechanisms which reduce information overload while at the same time meeting the a number of key goals. These are first, ensuring individual autonomy of action as much as possible; second, maintaining a power balance between “speakers” and “listeners” (many communication systems, electronic mail for example, are biased towards speakers); third, minimising hard-wired constraints and replacing them with a model of increasing effort and fourth, starting with support for social mingling and adding more formal mechanisms later on if needed.

A further goal of this work has been to try to describe these mechanisms in terms of a more formal theory which could then be applied to a variety of collaborative systems. The result of this work has been a so-called “spatial model” of interaction.

3.1. Key concepts in the spatial model

The spatial model, as its name suggests, uses spatial characteristics of virtual rooms as the basis for mediating interaction. In the model, a room becomes a space and each person therefore has properties of position and orientation within the room. People can move around the room relative to each other and can also turn to face each other. We identify two key concepts which allow people to govern their interaction.

FOCUS - represents a sub-space within which a person focuses their attention. A person is more aware of objects inside their focus and less aware of objects outside of their focus.

AURA - represents a sub-space within which a person projects their presence. An object within your aura is more aware of you than an object outside of your aura. We would like to stress that the concept of aura first originated in the Multi-G project at the Swedish institute of Computer Science [Fahlen 92].

Together, focus and aura contribute to your **AWARENESS** of other people. It is this awareness level that dictates the information that you perceive. Focus and aura are manipulated implicitly by moving around the room and through orientation. They may also be manipulated explicitly (e.g. by focusing in on a particular object). The use of focus and aura in this way imply that each person autonomously controls their awareness of others; that levels of awareness between two people need not be symmetrical; that just because you are speaking to me doesn’t mean that I have to listen to you and that I can’t stop you speaking (but I can stop listening to you).

In order to support communication scenarios beyond mingling, we can introduce various computer tools into rooms to manipulate focus and aura. We call these *focus/aura adaptors*. For example, a microphone tool might enhance my aura thus allowing me to project over many people and a table tool might fold together the foci and auras of several people into a semi-permeable sub-space within a room. Room shape might play a similar role, for example, an alcove might act like the table tool.

Focus, aura, awareness and adaptors are then the basic concepts that allow people to mediate interaction in large rooms. The following section outlines a more rigorous definition of these concepts in mathematical terms.

3.2. A Formal Model of Awareness, Focus and Aura

The concepts of focus, aura and awareness provide a means of mediating interaction which is intended to be useful across extremely diverse application domains. In order to realize this generality we propose a formal model which provides a framework for describing them in a consistent, application independent manner. This requires the introduction of one further concept, that of a communication medium. We assume that any room-based implementation will utilise one or more media for communication (e.g. sound, light or text). A model is required which will answer the following question:

"In a given room how aware is entity i of entity j via medium k?"

The Model

It is both convenient and useful to represent the factors which influence awareness as mathematical functions. This affords us both the flexibility of definition we require as well as a method to manipulate awareness in the form of a measurable quantity.

Consider an arbitrary room in which communication takes place between *n* entities via *m* media. The concepts of focus and aura are the most fundamental to the model. All entities define these quantities for each of the *m* media. We denote functions to represent them for entity *i* in medium *k* as f_{ik} and a_{ik}. Note that an entity may employ different focus and aura functions for different media. These functions evaluate to a measure of the focus (or aura) of a given entity on (at) another depending on certain *properties* of both of the entities concerned, for example their positions and orientations. Thus

$$f_{ik}: T_k \text{x} T_k \rightarrow R$$
$$a_{ik}: T_k \text{x} T_k \rightarrow R$$

Where f_{ik} is the focus function and a_{ik} is the aura function both of which map property values into the set of real numbers *R*. These properties are drawn from a set T_k for an entity *i* by the function t_{ik}. A measure of the focus of entity *i* on entity *j* in medium *k* is therefore given by:

$$f_{ik}(t_{ik}, t_{jk})$$

The form of the functions and the properties used is a matter for a particular application of the model and may be chosen according to the media involved and the intended sophistication of the end system. In spite of this, it is worth noting that we would expect T_k to often involve the position and orientation of various entities in the space (i.e. your focus and aura depend upon where you are standing and facing). More sophisticated systems might also involve other properties such as focal lengths and even degree of activity. The focus and aura functions are used to describe how an entity interacts in a particular medium, hence they are grouped to form a 2-tuple known as a medium description. We define a medium description D_{ik} for entity *i* in media *k* as follows:

$$D_{ik} = (f_{ik}, a_{ik})$$

This represents how a particular entity influences its interaction in a given medium. We can extend this further to define an entity description, E_i as the combination of media descriptions across all media in a room.

$$E_i = (D_{i1}, D_{i2}, \ldots, D_{im})$$

Now we consider how to represent awareness in a given medium. Awareness in medium *k* is itself a function that combines focus and aura functions.

$$A_k(f_{ik}, a_{jk}) : R^2 \rightarrow R$$

A_k is the function which evaluates to a measure of awareness of a given entity *i* to another *j* based on values of the focus of entity *i* (f_{ik}) on *j* and the aura of entity *j* (a_{jk}) at *i*. We are now equipped to answer the question:

"In a given room how aware is entity i of entity j via medium k?"

$$\text{Level of Awareness} = A_k(f_{ik}(t_{ik}, t_{jk}), a_{jk}(t_{jk}, t_{ik}))$$

It is worth emphasising that the awareness that entity *i* has of *j* is a combination of *i*'s focus on *j* and *j*'s aura at *i*.

Next, we model the tools which serve to amplify focus and aura in a room. A tool is itself an entity and so has an entity description specifying a set of focus and aura functions for different media. Upon selecting a tool a person acquires the awareness properties of the tool instead of their own. In other words, they replace their entity description with that of the tool they are using.

Applying the model

Our model allows the awareness that one person has of another to be calculated as a numeric quantity. This measure of awareness then needs to be interpreted in some concrete way within an end system. The process of interpretation lies outside the domain of responsibility of the model and will typically be part of the functionality of a *User Agent* (UA). In other words, a user agent presents an interpretation of awareness which assigns a real world meaning to the numerical measure. This may vary according to various parameters such as the media involved and the capabilities of a UA presenting the model. For example, in the sound medium a particular UA may interpret the level of awareness as a measure of volume whereas in a text medium it may use it decide whether to display a given message or not. To apply our model to the design of an end system, several steps therefore need to be followed.

- 1. Identify the media involved.
- 2. Identify the elements of the property set for each medium.
- 3. Specify focus and aura functions for entities representing people in each medium.
- 4. Specify awareness functions for each medium.
- 5. Specify additional focus and aura functions for any tools for each medium.
- 6. Decide how awareness values are to be interpreted by the system for each medium.

In the following section we follow these steps with a simple example, applying the model to the design of a text conferencing system.

4. Example Applications of the Model

4.1. A First example - applying the model to a text conferencing system

First we consider how the spatial model might be applied to a text conferencing system. It is important to remember throughout that this represents only one of possibly many applications of the spatial model.

Considering step 1, we have one medium, that of text messages. For step 2, we define rooms to be two dimensional spaces and let the properties of position and orientation (2-D vectors) define the domains of the focus and aura functions (i.e. we calculate focus and aura values from the positions and orientations of the people involved). Basic focus and aura might then be defined in terms of areas of the room floor shaped like segments of a circle. For any object in this area my focus/aura value would be 1. For any object outside, it would be 0. Thus, we have discrete functions where I am either completely focusing/auraing on an object or am not focusing/auraing on it at all. An example function along with its mathematical definition is shown in figure 1.

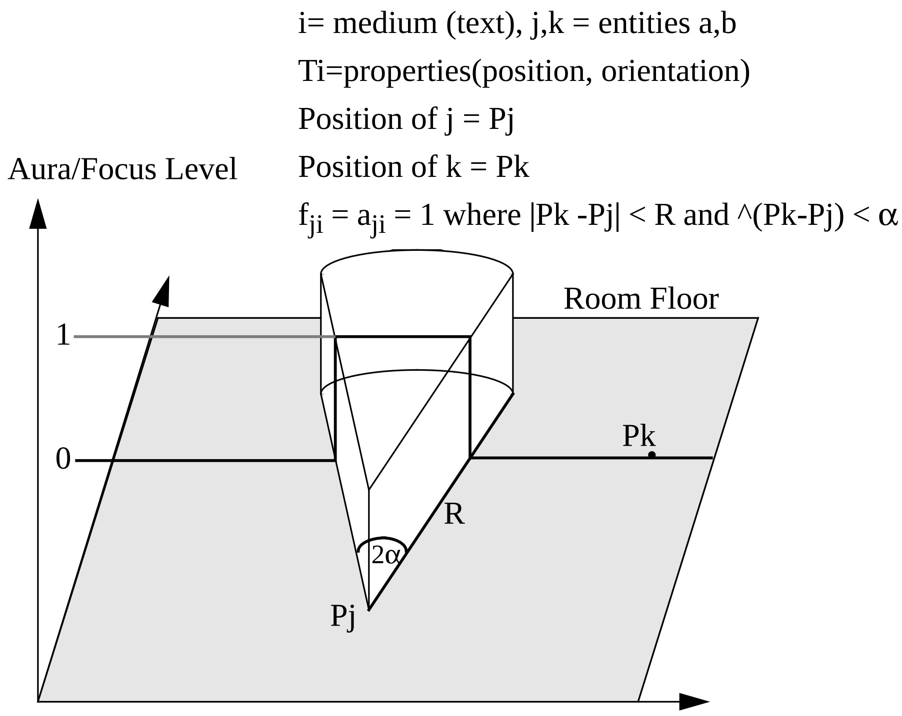

Figure 1: A Basic Function for Focus and Aura

I might manipulate these functions implicitly by moving around the room or by turning about. I might also manipulate them more explicitly (e.g. by focusing in or out). For example, focusing in would mean increasing the range of focus and decreasing the angle (so the focal area remained the same).

Under step 3, we can define an awareness function which simply adds together the relevant focus and aura values. More specifically, the awareness that entity a has of entity b is calculated as the sum of a's focus *as measured at the position of b* and b's aura *as measured at the position of a*. Thus, $A = f_{ik} + a_{jk}$. The result of this function will always be one of three possible values for A's awareness of B.

- 0 - when A is not within B's aura and B is not within A's focus;
- 1 - when A is within B's aura and B is not within A's focus or A is not within B's aura and B is within A's focus;
- 2 - when A is within B's aura and B is within A's focus.

Before we discuss various adaptors which might be introduced, we will briefly consider step 5, interpretation. One possibility might be the following. An awareness level of 0 implies *peripheral awareness* - you are aware that the other person is present in the room and you know where they are standing. This suggests that you are also informed of their movement. An awareness level of 1 implies *semi-awareness* meaning that you are also notified when ever they communicate. However, you don't hear what they say. Finally, an awareness level of 2 implies you actually hear what they say (i.e. receive their messages).

In terms of implementation, this might be represented by three windows at an interface: a map window showing the positions of everyone in the room; a notify window informing you of communication acts taking place and a speech window where you receive messages. The implementation would then let you move around the room in order to get close to people so that you could speak to them and hear them. Notice that, even with these basic discrete functions, some interesting properties emerge. In particular, I can choose to ignore you even if you are standing next to me shouting. In this case, I will not receive your messages but will still be informed that you are communicating (in a separate window). In other words, you can try to interrupt me, but I can separate your interruptions from my current conversation. In terms of power and control, this seems to have a pleasing sense of symmetry.

Now we go on to briefly consider the introduction of some focus and aura adaptors. Consider a more formal scenario, say a lecture or presentation, where a speaker wishes to address a large crowd. We can model this by replacing the speaker's current aura function with one which spans the entire room. In the simplest case this is a function which delivers the value 1 everywhere (see figure 2). This might be triggered by the speaker picking up an adaptor tool analogous to a virtual microphone. Notice that people can still ignore what the speaker is saying although they will always at least be notified that they are speaking.

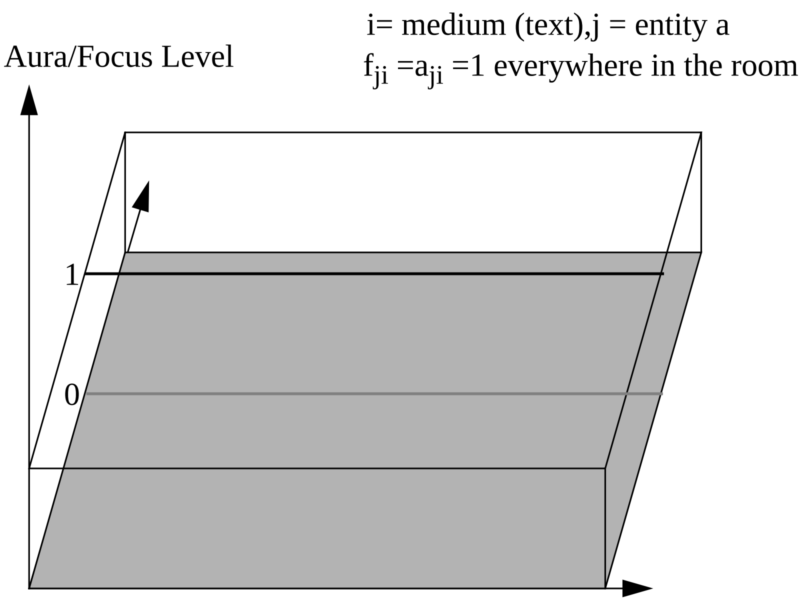

Figure 2: An Aura Function Representing a Virtual Microphone

Finally, consider the case where a group of people wish to form a sub-group within the room in order to have a (semi)private discussion. We assume that they don’t want an entirely isolated discussion which could be achieved by moving to a separate private room, but instead wish to maintain some degree of awareness of what is happening around them. We can realise this scenario through a table adaptor tool. When a person sits at the table their focus and aura are replaced by a common focus/aura function relative to the centre of the table (see figure 3). In effect, the focus and aura of all people around the table is folded into a common space. The implications of this are that the details of the conversation can only be perceived by those sitting at the table but that others are aware that there is a conversation taking place and that the sub-group are equally aware of people elsewhere in the room.

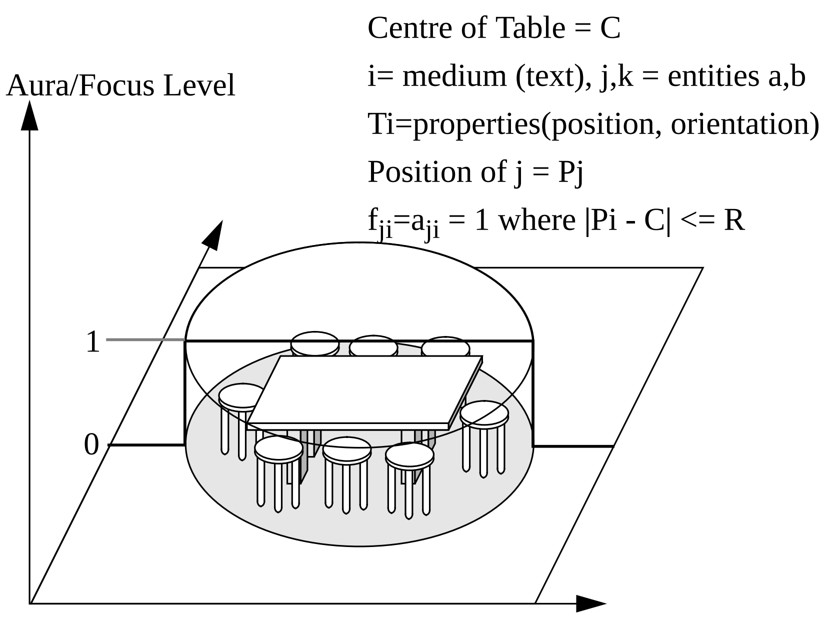

Figure 3: Functions Representing a Virtual Table

In summary, we have seen how the formal model of awareness, focus and aura might be applied to a text conferencing system by following each of the five steps mentioned above. Even at this relatively simple level, it should be possible to see the difference between the spatial metaphor and more explicit forms of floor control.

4.2. Further example applications

The spatial model is intended to be applicable across a diverse range of technologies. This section briefly speculates on some further applications. First it is interesting to speculate on the use of continuous as opposed to discrete mathematical functions for focus and aura to give a more fluid and gradual approach to awareness. Imagine a more sophisticated audio-conferencing system. In this case, we could choose functions which delivered more continuous values of focus and aura. A good function might have some kind of local maximum representing the peak of focus and aura, with its value asymptotically decreasing towards the edge of the room (e.g. a suitable quadratic function or normal distribution). Figure 4 shows the focus of one person and the aura of a second along one dimension of the room using such a function. As before, we can calculate the awareness that A has of B by summing the value of A’s focus at B and B’s aura at A to get some real number. However, in this case we have a continuous range of possible values as opposed to just three levels of awareness. These values might be interpreted in several ways. For example, they might be used to control the value of an audio channel such that as people moved towards you they would gradually become louder as you turned away they would become quieter.

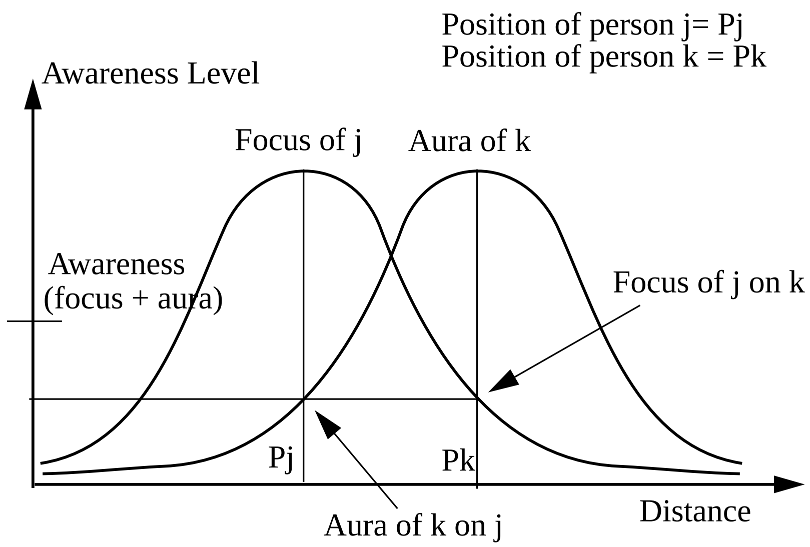

Figure 4: Continuous Functions for Focus and Aura

So far we have discussed the application of the spatial model to virtual worlds based on text or audio media. There are many other possible applications of the model. The concepts of focus and aura naturally apply to three dimensional virtual reality systems where they could be constructed as (probably invisible) volumes around objects. Indeed, pioneering work modelling 3-D aura has already been demonstrated by the Multi-G project. Furthermore, existing VR implementations often model a bounding space around an object which defines its sphere of interaction (e.g. for detecting impacts and collisions).To go a stage further, the virtual world and spatial models can be applied to any kind of system where a spatial metric can be identified. For example, we can measure distance between people in shared hyper-media systems in terms of number of links between them. The spatial model could then allow people to become more aware of each other as they moved closer within a hyper-media document (e.g. automatically opening up conferencing facilities if they came within a certain distance). An even more radical application might be to classify interests in a bulletin board system along a spatial dimension. In this case, the model would allow you to focus on specific topics and so become aware of other people with similar interests

Finally, note that the model need not only support interaction between people, it could also support interaction between people and objects. For example, as I approach a notice board in the virtual world it might become more aware of me. As a result it might redisplay itself to give a more detailed view of its notices.

5. CyCO - a distributed implementation

So far we have presented a number of models for building virtual worlds. This section briefly describes our current demonstration system. Our implementation is called CyCo, *Cyberspace for Cooperation,* and at present consists of a set of related demonstrators grouped around the theme of a rooms based text conferencing system (the advantage of text is that our demonstrator becomes easily shareable by many people across large networks such as the internet using readily available technology). The core of CyCo is a distributed environment of rooms through which people can move and within which they can hold conversations through the medium of real-time text messages. The system maintains room descriptions and is able to notify occupants of the comings and goings of other people as they wander about the world. In the most basic implementation, the spatial model is reflected through support for three different kinds of interaction: talking addresses the occupants of a room; whispering is directed at an individual but other occupants may be aware that the whisper occurred; shouting may be transmitted across several rooms CyCo is implemented on top of the ANSA Distributed Processing Platform [ANSA 89] and currently supports two user interfaces, an X Windows interface using the Motif widget set and a teletype interface based on the UNIX Curses C library. Figure5 shows an example screen from the X Windows interface. CyCo can also be configure to run an arbitrarily structured world of connected rooms and comes complete with an inbuilt mapper tool.

6. Summary

Our paper has presented a spatial model of group interaction within a distributed cyberspace of connected virtual rooms. The primary goal of this model has been to support cooperation on a *large scale*. The model itself aims to provide generic mechanisms for promoting and managing interaction which could be applied across a diverse range of systems from text conferencing to virtual reality. In order to illustrate the model we have included discussions of several possible applications including our current demonstration system called CyCo. We have then extended this metaphor to include the concepts of awareness, focus and aura as affordances for conversation management.

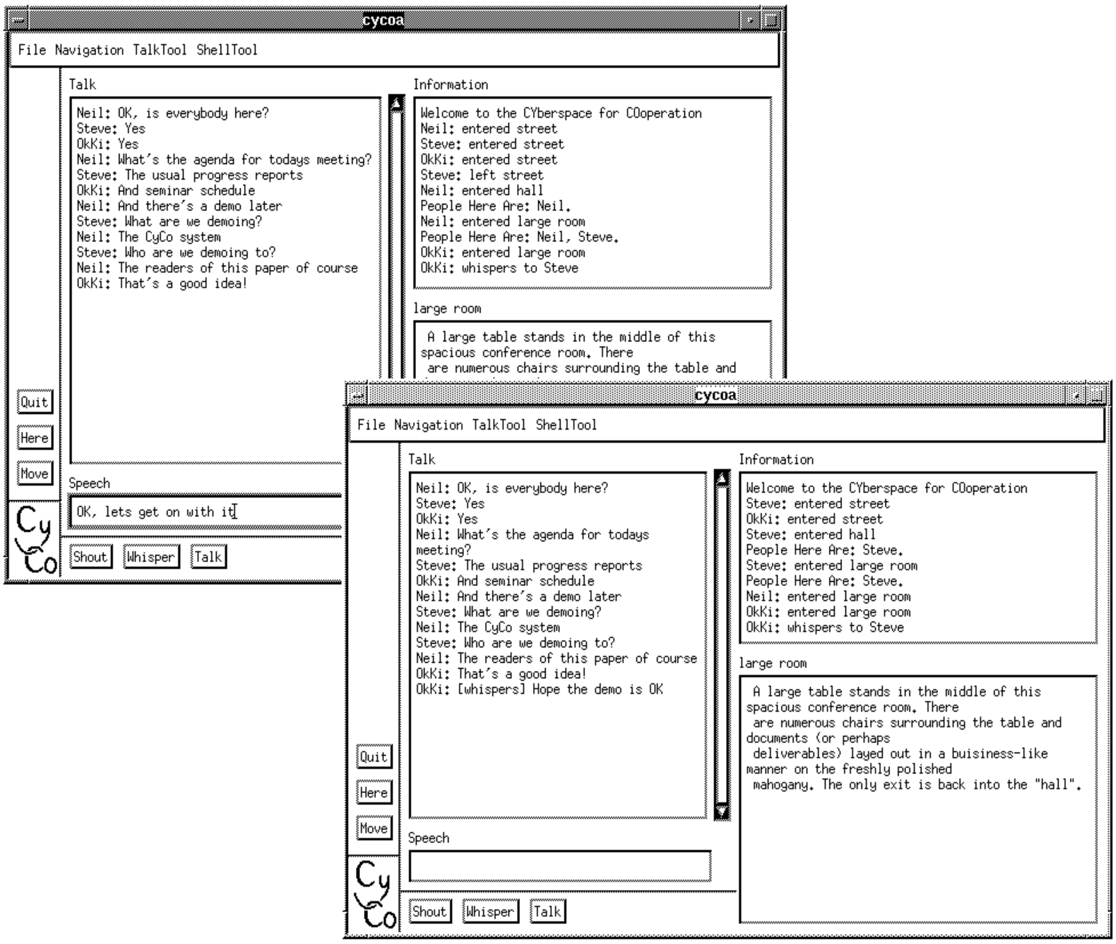

Figure 5: The Fundamental CyCo Interface

In conclusion, we believe that the virtual worlds metaphor is appropriate to the design of large scale collaborative systems. In fact, to go a step further, we believe that the metaphor is a sensible way to approach interfaces to distributed systems in general. We hope that the models that we have proposed can be applied across a wide range of systems to enable them scale effectively. Our long term goal would be to move towards a globally connected cyberspace rich in communication potential. Perhaps in a few years it may be possible to step through the window of your computer into the densely inhabited virtual city that is the global computer network.

References

[ANSA 89] *The ANSA Reference Manual*, Architecture Projects Management Ltd., Cambridge, UK, 1989.

[Benedikt 91] Michael Benedikt, *Cyberspace: Some Proposals,* in Cyberspace: First Steps, Michael Benedikt (ed), MIT Press, 1991, pp 273-302.

[Borning 91] A. Borning and M. Travers, *Two Approaches to Casual Interaction over Computer and Video Networks,* Proceedings of CHI'91, New Orleans, April 27-May 2, 1991, pp13-19.

[Bignoli 91] C. Bignoli and C. Simone, AI Techniques to Support Human to Human Communication in CHAOS, in Studies in Computer Supported Cooperative Work: Theory, Practice and Design, J. Bowers and S. Benford (eds), Elsevier Science Publishers, 1991.

[Bowers 88] J. Bowers and J Churcher, *Local and Global Structuring of Computer Mediated Communication: Dveloping Linguistic Perspectives on CSCW in COSMOS,* Proc. CSCW'88, Portland, Oregon, Spetember 26-28 1988.

[Clarkson 91] Mark A. Clarkson, *An Easier Interface*, BYTE, February 1991, pp227-282.

[Cohen 91] Michael Cohen and Lester F. Ludwig, *Multidimensional Audio Window Management*, in Computer Supported Cooperative Work and Groupware (Saul Greenberg ed), Hracourt Brace Jovanovich, 1991.

[Cook 91] S. Cook and G. Birch, *Modelling Groupware in the Electronic Office*, International Journal of Man Machine Studies, Volume 34, Number 2, pp 369-394, March 1991.

[Fahlen 92] Lennart Fahlen and Charles Brown, *The Use of a 3D Aura Metaphor for Compter Based Conferencing and Teleworking*, Porceedings of the 4th Multi-G Workshop, Stockholm-Kista, May 1992, pp 69-74.

.[Furnas 86] Furnas, *Generalised Fisheye Views*, Porceedings of the ACM SIGCHI'86 Conference on Human Factors in Computing Systems, 1986, pp16-23.

[Henderson 85] Henderson and Card, *Rooms: The Use of Multiple Virtual Workspaces to Reduce Space Contention*, ACM Transactions on Graphics, Vol. 5, No. 3, July 1985.

[Leevers 92] David Leevers, *Prototyping Multimedia Communications for Manufacturing SME's*, Presented at the CI Europe Seminar, 2-3rd July 1992 and also an internal report of BICC Central Development, Hemel Hempstead, UK.

[Lovstrand 91] Lennart Lovstrand, *Being Selectively Aware with the Khronika System*, Proceedings of ECSCW'91, Amsterdam, Hollandm Kluwer Academic Publishers, pp265-278.

[Morningstar 91] Chip Morningstar and F. Randall Farmer, *The Lessons of Lucasfilm's Habitat*, in Cyberspace: First Steps, Michael Benedikt (ed), MIT Press, 1991, pp 273-302.

[Pankoke 89] Uta Pankoke-Babatz (ed), *Computer Based Group Communication - The Amigo Activity Model*, Ellis-Horwood, 1989.

[Root 88] R.W. Root, *Design of a Multi-Media Vehicle for Social Browsing*, Proceedings of CSCW'88, Portland, Oregon, Spetember 26-28 1988, pp25-38.

[Sarin 91] Sunil Sarin and Irene Greif, *Computer-Based Real-Time Conferencing Systems,* in Computer Supported Cooperative Work: A Book of Readings, Irene Greiff (ed), Morgan Kaufmann, 1988, pp397-421.

[Sarkar 92] Manojit Sarkar and Marc H. Brown, *Graphical Fisheye Views of Graphs*, Proceedings of ACM SIGCHI'92 Conference on Human Factors in Computing Systems,, May 3-7 1992, pp 83-91.

[Smith 92] Jennifer Smith (ed), *Frequently Asked Questions: Baisc Information about MUDs and MUDing,* Posting to the alt.mud USENET News Group, July 1992.

[Victor 91] F. Victor and E. Sommer, *Supporting the Design of Office Procedures in the Domino System,* in Studies in Computer Supported Cooperative Work: Theory, Practice and Design, J. Bowers and S. Benford (eds), Elsevier Science Publishers, 1991.

[Winnograd 86] T. Winnograd and F. Flores, *Understanding Computers and Cognition: A New Foundation for Design*, Norwiid, New Jersey, Ablex 1986 and Addison-Wesley 1987.

A virtual team (also known as a geographically dispersed team, distributed team, or remote team) usually refers to a group of individuals who work together from different geographic locations and rely on communication technology such as email, FAX, and video or voice conferencing services in order to collaborate. The term can also refer to groups or teams that work together asynchronously or across organizational levels. Powell, Piccoli and Ives (2004) define virtual teams as "groups of geographically