Chapter 3 Collaboration Theories

"England were bewildered, yes, and bothered and bewildered, by a side of soccer sorcerers who, at times, seemed capable of reading each other's minds. This was the nearest thing to telepathy on a football field, and I doubt if its like will ever be reproduced in my lifetime."

'Finney' (1960), on England vs. Hungary (1-7)

Budapest, in: Taylor and Jamrich (1997).

3 Collaboration Theories

This chapter consists of a review of key features of human-human communication and collaboration. The information is divided into four sections: verbal communication issues, non-verbal communication issues, meta-collaboration issues, and a comparison of these issues to the types of mediated collaboration technologies discussed in Chapter 2.

3.1 Introduction

This chapter is a review of psychological and sociological knowledge about human-human communication and interaction for collaboration in the real world. It places real world collaboration in the context of CVE type cooperative work, and the meta-collaboration that starts of the task process of collaborative work. It is by no means exhaustive in describing all types of human-human communication and interaction, but rather considers in pertinent detail each of the issues relevant to collaborative work. This approach has been adopted in order to clearly identify the type of collaboration support that is needed, and the limitations of CVEs in the support of these activities, at the time this thesis was produced. Furthermore, it compares each type of synchronous distributed groupware (described in Chapter 2, section 2.3.1), to those demands that collaborative work would put on them. The aim of this chapter is to give an overview of the issues involved in meta-collaboration, collaboration, and mediated collaboration, describe the key features of those types of communications and interactions in order to identify the type of collaboration support that is needed to allow users to collaborate in a CVE, and finally, to discuss how CVE technology

compares to other synchronous distributed groupware technologies, in terms of their potential to support collaborative work, in theory and in practise.

The next section presents an introduction to the key strands of human-human interaction and the related CVE collaboration issues (section 3.2). Next, a comparison is made between these key strands and other mediated collaboration technologies (section 3.3). Finally, some conclusions are drawn from this review (section 3.4).

3.2 Review of Collaboration

Human collaboration depends both on complex social interaction and the physical medium in which these interactions take place. Human social interaction depends on the shared definition of a situation that organizes and governs people's involvement in the goings-on, the type of interactions that take place, and the order of the interactions and social events that take place. This shared definition of the situation is constantly modified based on the interaction negotiations, which take place in the activity space. People subtly (or not so subtly) employ spoken word, gesture, time, movement, position, orientation and objects in a self-invented, group-maintained social ritual which could be termed meta-collaboration: the construction of the social reality par excellence. "We walk into a situation, identify its features, and match our actions to it." (Suchman, 1987). Within this social reality takes place the actual collaboration CVE technology aims to support: goal-directed social activities that people perform when they work jointly with others on a certain task. However, these collaborative actions only become significant and understandable to the other collaborators in relationship to their publicly available, collaboratively organised world of artefacts and actions (Boden, 1994, Hindmarch, 1997). Thus CVEs need to support the

complete interaction process that leads to collaborative work. The analysis of the social activities involved in small group collaboration, presented in this thesis, therefore also takes into account the objects used in the interaction, the primary physical medium in which the interactions take place, the CVE, and to some extent the secondary physical medium, the real environment of each user.

The pervasive attitude in face-to-face interaction analysis, is that the many strands of communication and the communicative function of each one strand, is better understood in the context of the operation of the other strands than in isolation. All the different communicative strands - speech, gesture, posture, body movements, orientation, proximity, eye contact and facial expressions - should be thought of as woven together to form the fabric of conversation, and the particular texture of interaction can only be understand properly, by seeing the relationship of the different strands. As evoked above, collaborative actions are similarly, only significant in relationship to the other collaborations and collaborators inside the collaborative organised world of artefacts and interactions, and the success of the collaboration depends on how effectively the interactions are progressing during the collaboration process. The acts of collaboration are triggered by previous acts of collaboration, either our own or those of others with whom one is collaborating, and they are also triggered by changes in the collaborative space, be it objects in that space or other collaborators, through, speech or movements of some kind. The analysis of the social activities involved in small group collaboration, presented in this thesis, therefore also takes into account the sequence of collaborative activities, and the analyses differences between the different stages of the collaboration process (see Chapter 8, section 8.3).

Alternation of action and inaction, and the subsequent rotation of performance among individuals in a group is the most salient feature of group dynamics. In order to clarify the social activities involved in collaboration, the author of this thesis has included all aspects of the social ritual in the analysis, including some elements of social conduct such as for instance dress, touch, and facial expressions, which are either not or only rudimentary supported in the CVEs under research in this thesis (see section 3.3). Not all aspects of interaction are covered in the same depth, however, each strand is discussed to assess how these activities would be supported by the CVE, and how the absence of support could influence the usability of CVEs as a collaboration medium. Table 3.1 lists the elements of social conduct and their respective definitions discussed in this chapter.

Social Behaviour	Definition
Verbal communication	The exchange of audio information to establish and maintain contact between individuals engaged in focused or unfocussed interaction.
Phatic communication	The exchange of stereotyped phrases and commonplace remarks to establish and maintain a feeling of social solidarity and well-being.
Spatial regulation	The arrangement of single or combined average body-size related spaces around and between people and objects, signifying temporary or permanent micro-territories; where each cultural tradition has its own micro-territorial sizes and arrangements.
Proxemic shifts	Patterns of interpersonal distance in face-to-face encounters accompanying and influencing changes in the topic or in the social relationship between speakers (i.e. situational shifts).
Turn taking	Non-verbal communication accompanying verbal communication has an important role in the understanding of social interaction and turn taking during interaction. Common turn taking cues are head nodding, face looking, smiling, head touching, and speaking, including simultaneous speech.
Peripheral Awareness	A subtle monitoring of the other participants' activities; the individual monitors the other participants' activities without getting involved. This is largely accomplished through alignment towards the focal area of activity, such as a document.
Trust Building	Establish and confirm ones perceived trustworthiness as a competent collaborator, by

	being perceived by the other participants as acting according to the social norms.
Reciprocity	An individual's ability to be simultaneously both perceiver and perceived of their own embodied actions as well as the perceiver of others' actions.
Indexicality	Our ability to point at objects and locations and refer to them with indexical expressions such as 'that', 'there', etc.
Gaze	Gaze direction, gaze duration, gaze patterns, gaze awareness, mutual gaze and head turning indicate direction and type of attention given to something, aid in turn taking, and provide communication feedback.

Table 3.1: Elements of Social Conduct during Human-Human Interaction.

3.2.1 Verbal Communication

The universality of languages is partly pure syntax, and partly based on the way in which humans everywhere perceive and think about the world. Verbal communication has two aspects: language production and language comprehension.

Verbal communication in the CVE is mediated by the audio connections between participants. The quality of the transmission of the audio signal depends partly on the audio settings, and the microphone, the network congestion, and partly on how effective the human user is utilising the audio channel to communicate in the CVE.

3.2.2 Non Verbal Communication

Nonverbal communication is communication by vocal quality, facial expression, position of and change in position of hands, feet, and the rest of the body, in relation to the space and contents around. The nonverbal channels are continuously attended to, and do communicate information – primarily affective in quality and connected with personal perception, attitude and relationships. They are generally used to support verbal communication, but they are not as easily controlled as verbal communication and can therefore sometimes be downright contradictory.

3.2.3 Meta Collaboration

There are several stages during the interaction process that are crucial for the success of collaboration. The author of this thesis has attempted to summarise this process in figure 3.1.

```
┌─────────────────────────────────┐
│   ┌──────────────────────────┐  │
└──▶│   Acknowledge presence   │  │
    └──────────────────────────┘  │
                 │                 │
                 ▼                 │
    ┌──────────────────────────┐  │
    │   Navigate into position │  │
    └──────────────────────────┘  │
                 │                 │
                 ▼                 │
    ┌──────────────────────────┐  │
    │   Contribute to meta-    │  │
    │        planning          │  │
    └──────────────────────────┘  │
                 │                 │
                 ▼                 │
    ┌──────────────────────────┐  │
    │        Negotiate         │  │
    └──────────────────────────┘  │
                 │                 │
                 ▼                 │
    ┌──────────────────────────┐  │
    │  Contribute to task at hand │
    └──────────────────────────┘  │
                 │                 │
                 ▼                 │
    ┌──────────────────────────┐  │
    │  Acknowledge utterances  │──┘
    └──────────────────────────┘
```
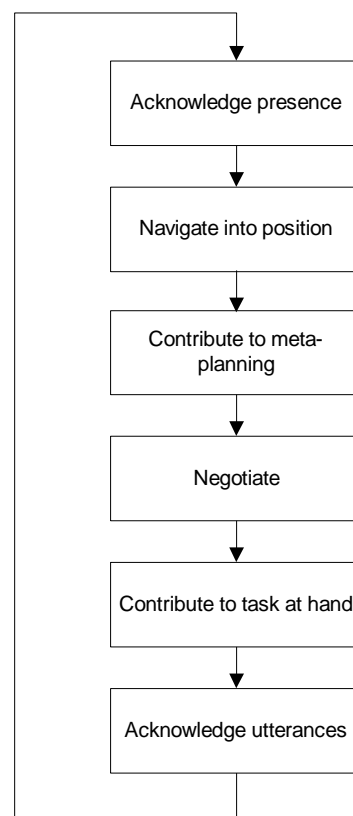
Figure 3.1: Interaction-flow in a (CVE) collaboration task.

People have to perceive each other, acknowledge each other's presence, they have to fine-tune their position in the space and with regards to the other people and the shared objects in the space, assume roles in the actual collaborative task, and participate accordingly, collaborate towards their shared goal, and switch from the collaboration proper to a more individualistic role. In order to do this accurately they

ideally have to be able to see and be seen, make sounds and be heard and hear others, touch and be touched, reorient perspective, etc. This intertwining of mutual perception is called reciprocity (see section 3.2.7). An individual has to be both perceiver and perceived, of their own actions as well as others' actions. This monitoring of the goings-on in the shared space is called peripheral awareness (see section 3.2.8).

3.2.4 Phatic Communication

Phatic communication (Kendon, et al, 1975)serves to establish and maintain a feeling of social solidarity and well being, by exchanging words. Phatic communication can be usefully described as a ceremony functioning as a rite of passage, easing and signalling the transitions to and from conversational interactions. The transactions during arrival and departure are ceremonial transition from the broader social macrocosm to the momentary microcosm of the encounter, and also from the recent shared microcosm to the readopted macrocosm. Transitional ceremonies in the opening phase can be called proleptic, and transitional ceremonies in the closing phase can be called analeptic rites (see figure 3.2).

The prime function of phatic communication is the communication of facts about speakers' identities, attributes, and attitudes, and these facts constrain the nature of each particular interaction. Managing the behavioural resources of phatic communication is a basic skill essential to a major part of the psychosocial transactions that make up daily life. Phatic communication and associated behaviour are very important social and psychological instruments, in that the cumulative consensus about the relationship reached as a result of repeated encounters between the two participants constitutes the essence of that relationship. Phatic communication

applies to choices from a limited set of stereotyped phrases of greeting, parting, commonplace remarks about the weather, and small talk, and yet it has an important function in helping the participants to reach what Goffman (1967) has called the "working consensus" of the interaction.

```
┌─────────────────────────────────────────┐
│ Meta Collaboration                       │
│   ┌───────────────────────────────┐      │
│   │                               │      │
│   │     Begin CVE Interaction     │      │
│   │       Proleptic Rites         │      │
│   │                               │      │
│   └───────────────────────────────┘      │
│                  │                        │
│                  ▼                        │
│   ┌───────────────────────────────┐      │
│   │                               │      │
│   │     Collaboration Proper      │      │
│   │                               │      │
│   └───────────────────────────────┘      │
│                  │                        │
│                  ▼                        │
│   ┌───────────────────────────────┐      │
│   │                               │      │
│   │      End CVE Interaction      │      │
│   │        Analeptic Rites        │      │
│   │                               │      │
│   └───────────────────────────────┘      │
└─────────────────────────────────────────┘
```
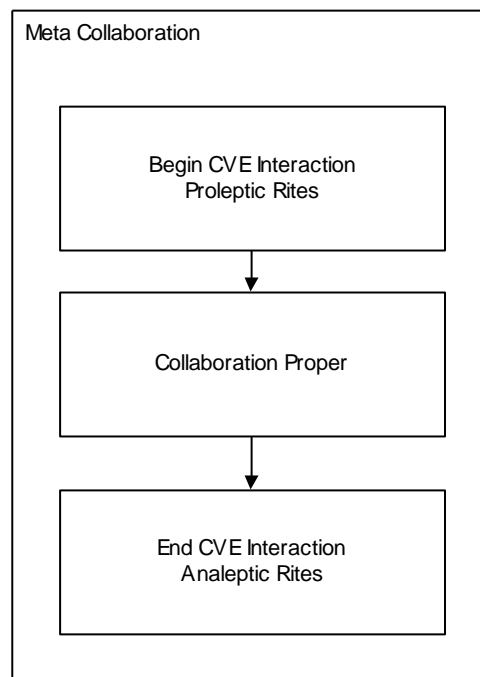
Figure 3.2: Three stages of collaboration.

The temporal structure of interactions is divided into three major phases, the opening phase, the middle phase, and the closing phase. The function of the behavioural activities that characterise the opening phase is to lubricate the transition from noninteraction to interaction, and to ease the potentially awkward tension of the early moments of the encounter, before the main business of the encounter is embarked upon in the middle phase. The closing phase is once again a transitional phase, easing the transition from full interaction to departure. Phatic communication strongly characterises these marginal phases of communication.

The function of phatic communication during the opening phase of interaction is:

- to provide the participants with acceptable means of stating the outlines of the roles they are prepared to play in the oncoming interaction, at least in terms of status, psychological distance and territoriality;

- to extend and accept invitations to sociolinguistic solidarity, and

- to facilitate the comfortable initiation, free from tension and hostility, of the interaction.

The function of phatic communication in the closing phase of interactions is:

- to achieve a cooperative parting, in which any feelings of rejection by the person being left can be assuaged by appropriate reassurance of the person who is leaving, and

- to consolidate the relationship between the two participants, by means of behaviour which emphasises the enjoyable quality of the encounter, the mutual esteem in which the participants hold each other, the promise of a continuation of the relationship, the assertion of mutual solidarity, and the announcement of a continuing consensus for the shape of encounters in the future.

The sequence of events in a typical opening of an interaction (Kendon et al, 1975):

1) Eye contact. To accept eye contact is the first signal of acknowledgment that one accepts the other participant's invitation to engage in an encounter.

2) The exchange of "distant" gestures of greeting or acknowledgment. These, exchanged between acquaintances, are much more understated in our culture than our distant gestures for parting, involving only slight movements of the hand and arm, or head.

3) The participants assume an appropriate, conventional facial expression of cordiality, or polite attention, or merely of attention, depending on the previously established or anticipated relationship between the two participants.

4) The participants reach the appropriate proximity for the remainder of the opening phase of their interaction (see section 3.2.5).

5) The exchange of conventional contact gestures of greeting, as appropriate to the relationship between the participants.

6) The participants take up their mutual bodily orientation, in postures appropriate to their relationship.

7) The exchange of stereotyped linguistic symbols used as tokens in the transactions of phatic communication.

8) Indication by the participants that they would like to initiate the main business of the interaction, by the use of various signals of transition. These include such actions as an abrupt head movement, moving the head slightly upwards so as to allow the establishment of eye contact on a level gaze; a slight shift in posture, sometimes an adjustment of proximity, often slightly increasing the distance between the

participants, and sometimes a linguistic marker such as "Well..", or more overt transitional comments such as "What I came to see you about was...".

CVE interaction typically does not allow for eye contact (step one). Distant gestures are limited to calling, or being called by, another participant (step two). Cordiality (step three), can only be expressed verbally, due to the absence of facial expressions. This means that step one, two and three of the opening phase ritual can not take place, thus suggesting that initiating an interaction within the CVE needs extra support. This would suggest that a CVE interaction starts at step four or five; moving into proximity of another participant, and greeting that participant in order to signify an attempt to start interactions.

The closing phase of interaction (Kendon et al, 1975):

1) The initiator of the closing phase performs appropriate signals of transition, indicating their desire for the onset of the closing phase and the end of the middle phase. Transition signals can take place on the visual channel, with a sudden prolongation of the typical duration of eye contact, or avoidance of eye contact for a longer period than conventional during the middle phase, or a roaming gaze. Greater shifts in orientation or of posture are another transition signal, or overt glances, facial expressions where the momentary interactional state of the middle phase makes no call for such an indication of apparent cordial, attentive agreement. Also, linguistic signals of transition from

the middle phase include the same sort of utterances as are used in the opening phase, such as "Well...".

2) Display of behaviours that emphasise the departure, such as finishing a drink, or cigarette, folding up and putting away belongings, etc. Proximity is also manipulated, changes in orientation, frequent glances towards the anticipated direction of departure, increase in vigour of facial expressions, loudness, such as would be necessary for audibility and visibility at a greater distance.

3) Exchange of tokens of phatic communication.

4) Exchange of conventional contact gestures of parting, where and as appropriate to the relationship. (Fourth and fifth stages are usually simultaneous.)

5) Adoption of conventional facial expressions of parting, where and as appropriate to the relationship.

6) Increasing distance between the participants.

7) Exchange at an appropriate distance of conventional distant gestures of parting.

8) Termination of the encounter by breaking eye contact.

Conversations can be terminated amicably only by mutual consent. The initiator of the closing phase has to obtain the consent of the noninitiator through the process of offering gambits on various communicative channels, and the closing phase can be developed further only when the gambit has been seen to be accepted. The social process of parting normally has an essential feedback component. Only when the appropriate feedback has been received, can parting progress, otherwise continual re-

entry to earlier stages of the interaction is necessary. Conventional parting is thus a cooperative achievement obeying rather strict constraints.

CVE interaction during the closing phase can typically only make use of two out of these eight sequential events, which would mean, especially in the light of the fact that parting obeys strict rules, that ending CVE interactions may be more complicated and more likely to confuse participants somehow. The two types of interactions that are available during the CVE closing phase of interaction are: three and six; respectively the exchange of typical phatic phrases, and the increase in distance between the participants.

3.2.5 Proxemic Shifts

Proxemic shifts are shifts in face-to-face interaction (Erickson, 1975). Erickson found that changes in interpersonal distance during interaction, sometimes accompanied by changes in body orientation, seem to accompany changes in the topic or in the social relationship between speakers, so called situational shifts. On the basis of this picture of the whole, relationships among parts could be inferred. The data confirmed that proxemic shifts occur very frequently at the beginning and ending of segments of interaction that can also be identified by changes of speech content and style, and by changes in the interaction process. The types of proxemic shifts that are likely to occur are described in the section on phatic communication (3.2.4).

3.2.6 Spatial Regulation

Spatial regulation is governed by micro-territorial perceptions and attitudes of individuals occupying that space (Scheflen, 1975). Displays of spatial movements are

part of the basis for comprehending and measuring activities and relations between activities. Territory can be defined as an area of open space that is bounded for a time in some discernible way and used by an animal or a group of animals (human animals included). Micro-territories range from a few inches in diameter to a few square yards, and they have a wide range of duration. Spatially, these mini-territories are organised in levels. A number of very small ones are assembled in particular and conventional ways to form a slightly larger one, and these larger ones are in turn put together to form a territory that is larger still, and so on. A very small one is the "spot", a bigger one is the "cubit", the "k" space, the "location", the "module" and the nucleus. Each cultural tradition has its own micro-territorial sizes and arrangements.

Spot

A spot does not have a standard size. A spot is located on the physical surfaces of floors or furniture, but also on the human body itself and marks the boundaries of a territory. Under conditions of medium or low density other members of that community and culture will respect the claim of a spot. Traditional constraints on taking over, touching or looking at a spot are accompanied by larger proxemic or interpersonal spacing patterns, which seem culturally defined.

K-Space

The body is based in a larger space than a spot, and more exactly defined. It is referred to as the cubit, and represents the average width of the body from elbow to elbow. Larger areas of space are exact increments of the cubit. The minimal space allocation for a person in a stationary posture is four cubits; this is called the "k"

space. In high density, standing crowds, each participant can hold but a single "k"

space, but ordinarily he or she uses more space than this.

Cubit

Touch can be described in terms of cubit-sized elements. Two lovers in side-by-side

relation at close interpersonal distance occupy adjacent cubit-sized locations. The

tactile connection makes their cubits adjoined. By the same token gaze-holding

involves a relation of two faces at some distance or other. So this communicational

activity employs a channel or "k" space of cubits. Participants in communication do

not usually act as organismic wholes. Instead they use bodily regions differently. A

single person may be simultaneously engaged at one point in time in orientational

relationships to one side of him, in tactile relations to the other side, and in gaze and

vocal relations across the group. These differential activities employ spaces that are

channel-like in form and "k" sized. These spaces are treated as territories.

Location

The location is that increment of space that is ordinarily occupied by one person for

some finite interval of time. Territorially speaking the location is an integration of "k"

space cubit and spot sub territories. In normal situations a person is allotted four "k"

spaces in standing activities, a seated location is three cubits "k" spaces with extra

lateral spots adding up to four "k" spaces or sixteen cubits. The standing location is

highly fluid since it is not concretised into built objects and is thus susceptible to

considerable expansion and decrease under a variety of contextual conditions.

Location size varies with density, status, affiliation, and other parameters.

Modules

A row or a queue of locations is a module. A module can be almost any length. If there is plenty of room an array of people will space themselves in accordance with their relationships and thus mark off locations. Strangerhood will be demonstrated by turning away from each other a bit, by placing arms between themselves, and by avoiding gaze and conversation. These instances of the closing of the legs or arms or clasping hands in front of them is equivalent to leaving a cubit of space. One can touch or gaze or stand closer in showing or establishing an affiliation. The side-by-side module is often equipped or built into an item of furniture. People leave at least a cubit of space if conversation is held, but distances vary with cultural habits. In British-American culture a full location is left between conversants if they are standing. A three-location face-to-face "x" module is thus formed with an open location in the middle.

Nucleus

Another very common type of module has the locations placed at right angles to each other. The occupants are thus oriented at an angle of about ninety degrees, but they may turn their heads to interact in face-to-face relation. Often people form small triangles, squares, or circles. Two arrays face each other across a space or table. In these cases more than a single module of space is required. The central area of tables and seats are called the "nucleus" of the total area, and the location strip around this area is called the "region". These regions are used for passage and storage. The nucleus and region arrangement can be found in many cultures dates back a long time and all show a magnitude of the cubit arrangements.

Territories

Vine (1975) reviewed human territoriality compared to other species' territoriality. Humans often show strong attachments to areas they use frequently, and we may monopolise these core parts of our ranges to a moderate degree. Familiarity with an area through frequent use can apparently be one factor enhancing our dominance if others enter it. Territoriality in humans, even when reinforced by legitimated ownership of space or motivation to maintain privacy, is typically highly ritualised and frequently amicable, unless the intruder poses any genuine threat to the owner. For the most part we learn to avoid attempting to enter true territories. Dominant or higher status individuals can command larger personal spaces and larger equilibrium distances from others. There is some evidence that personal spacing is larger on familiar ground. Group spacing seems to be maintained by concerted adjustment of the bodies involved; other persons may detour behind them or apologise when forced to pass in front or through them and will act in other ways to attempt to avoid offering a threat or seeming to invade others' privacy, such as deliberate gaze-avoidance or distance greeting rituals as a prelude to actually interacting. When we are very closely crowded we seem to deal with the violation of our space limits by a psychological cut-off process that "depersonalises" others, so that we can then treat them as physical obstacles rather than as individuals. This will only be effective for short periods, and sustained face-to-face confrontation leads to a necessity to react and interact in one way or another.

How well the perception of micro territories translates from the real world to the CVE is not quite known, although it is to be expected that virtual body size and the size of virtual spaces, and objects may not be designed with the same ratio as their real world

counterparts. In what way this affects interaction, and whether and when it has positive or negative effects remains to be seen.

3.2.7 Reciprocity

An individual has to be simultaneously both perceiver and perceived of their own embodied actions as well as the perceiver of others' actions (Robertson, 1997). This intertwining of mutual perception is a necessary condition for direct collaboration. Hindmarch et al. (2000a) found that collaborators pause or vary the speed of approach towards the intended locus of attention in coordination with amendments to their indexical utterances (see section 3.2.9), while judging the degree of attention and understanding of their listeners from the feedback derived from their respective redirections of gaze, movements, and utterances, or lack thereof.

In CVEs it is possible to perceive without being perceived, for instance using the bird's-eye view. Deliberately showing non-reciprocating views of the CVE or objects in it to different participants may be a desired situation, as when multiple users are simultaneously querying the same database, whilst sharing the same space the database is visualised for all participants to allow collaboration (Snowdon, 1996). However, any unintentional non-reciprocating perceptions of the same space will contribute to a break down of the understanding of each other's perspective. Hindmarch, Fraser, Heath and Benford (1998) found that CVEs' narrow field of view creates a fragmentation of the perception of the goings on in the CVE, so that participants cannot make sense of talk and activity because they cannot easily look from the speaker and the objects referred to and back again. Improvements to the display of actions have been proposed by Hindmarch et.al. (1998), such as:

- Subtle highlighting of region of space where a user is looking, thus creating a visible display of each

- CVE participants view frustrum to the other participants. Also, displaying speech bubbles in the CVE near a speaker when audio is sent, and ears appearing or enlarging when audio has been received.

- Highlighting the target object or region a CVE participant is pointing at with a visible ray of light between arm and object.

- Extending both virtual arms when a CVE user selects and grabs an object.

- Showing the manipulated object in wire frame whilst it is moving.

3.2.8 Peripheral Awareness

An individual's ability to work on collaborative tasks relies upon peripheral awareness of the others and a subtle monitoring of the other participants' activities (Heath et. al, 1995). The organisation and coordination of focussed and unfocussed collaborative work is enabled by this ability to monitor the activities of others' present in the same space. Collaboration between people sharing the same workspace involves the ongoing and seamless transition between individual and collaborative tasks, shifting their attention from a state of peripheral awareness to focussed awareness. Collaboration can be broken down into unfocused collaboration, where the individual monitors the other participants' activities without getting involved, and focused collaboration, where individuals are closely working together. Both focused and unfocused collaboration are largely accomplished through alignment towards the focal area of activity, such as a shared object or current speaker, where individuals

coordinate their actions with others through peripheral monitoring of the others involvement in the activity at hand.

Hindmarch and Heath (2000a), researched how people working in a control room shift each others attention from peripheral to focussed by establishing mutual orientation towards shared objects. They found that the visual conduct of a participant is designed, with talk and speaker's orientation, to have another notice something within their environment, and to see it in a particular way. The results of their workplace studies show that movement, coupled with the speaker's talk and bodily orientation is designed to encourage another to determine and confront an object within the local milieu, and forms a critical resource for that object to be seen or located by the other. They conclude amongst other things, that technologies mediating synchronous collaboration show a relative weakness in their support for these resources to establish mutual orientation to objects of interest.

Peripheral awareness in CVEs is limited to the field of view and the happenings in that field of view. If a participant cannot oversee the complete space and the happenings sufficiently to form an understanding of the goings on, their peripheral awareness will be limited to the view of the graphical window of the CVE on their desktop and any sounds they can hear from the CVE. If a CVE employs spatialised sound, participants' peripheral awareness will be slightly improved. In order to allow participants a peripheral awareness of the goings on in the CVE space they can view. They need to be able to distinguish other participants, and their activities in the CVE, and the results of their own and other participants' actions upon the objects in the CVE.

MASSIVE (a CVE described in Chapter 2, section 2.5.2), uses an 'out of body camera view' onto the virtual world (Greenhalgh and Benford, 1995). This provides participants with a view onto the CVE that includes the their own torso in the foreground of the scene, thus increasing the field of view and making the virtual arm visible on screen (with the viewpoint situated through the embodiment's 'eyes' it was not possible to see the virtual arm point). Other attempts at increasing the size of the field of view are the use of peripheral lenses (Robertson, Czerwinski, and van Dantzich, 1997), two additional windows with views of the virtual environment to the left and right, distorted to allow more information to be rendered within the smaller horizontal space, and a peripheral glance feature (Hindmarch et.al., 1998), that allows the user to momentarily swap their focus of attention to the left or right peripheral view, with each peripheral view providing a vertically compressed view of the left and right in two inch graphical rendering on each side of the main CVE window. The peripheral side that is currently viewed is presented as the main window, and the others distorted.

Results from attempts to increase the field of view by using peripheral lenses has shown that the extended representations of action that are provided by the lenses help CVE users overcome problems observed in object-focused interaction to some extent (Fraser, Benford, Hindmarch and Heath, 1999). However, not all interaction problems are improved by the use of these lenses, and the lenses themselves introduce a distorted view of the environment which has been observed to cause confusion.

3.2.9 Indexicality and Deictic References

Movement, the speaker's talk and bodily orientation form the critical resource for an object or another intended locus of attention, to be seen or located by others (see section 3.2.8). The speaker's talk, at this time, consists of many indexical expressions referring to the intended locus of attention. Indexical expressions or indexicals are words like 'that', 'this', 'he', 'she', etc and their interpretation is very much dependent on the external context (the state of the world) and internal context (the state of the conversation) in which they are used. Communication about a task will refer to the artefacts used as part of that task, using a variety of movements, such as changes in bodily orientation and gestures.

Hindmarch and Heath (2000b) found that people interact by controlling the sense and significance of features of themselves and their surroundings in regard to others around them. For instance, a speaker's movements and changing bodily orientations are aimed to provide deictic references to support indexical expressions. Deictic reference or deixis refers to our ability to point at objects and locations and refer to them with indexical expressions such as 'that', 'there', etc. The deictic gesture is interactionally accomplished with regard to the concurrent conduct of the co-participants (Hindmarch et al, 2000a). Furthermore, because these social dynamics form a large part of our interactions, the effectiveness of the interaction would be expected to suffer greatly if deictic references cannot be used.

If CVE participants cannot see each other point than this kind of reference is impossible. If and when a CVE participant wants to attempt to change the state of the attention of other participants' from peripheral awareness to focussed collaboration,

they have to make 'recipient-designed references' to the intended locus of attention. If they are provided with a pointing device in the CVE that is visible to the other participants, they will able to guide other participants' view to precise locations in the CVE. Without a pointing device, CVE participants can only use indexical expressions to refer to the shared object or other place of focus, to which they are trying to draw other participants' attention.

Observational analysis of CVE participants interacting in a virtual moving task (Hindmarch, Fraser, Heath, Benford, Greenhalgh, 1998), revealed that participants do feel the need indeed to make heavy use of explicit indexical talk to illustrate their actions and describe their visual conduct, to compensate for any lack of visibility due to the small field of view, the lack of definition of remote occurrences, and the general lack of support for deictic references. Apparently, their CVE participants had problems assessing and monitoring the perspectives and activities of other participants even when the other participants' virtual embodiment is in view. When an indexical utterance is produced, individuals did not seem to be in an immediate position to see the speaker's virtual embodiment alongside the referenced object, which meant that participants were repeatedly observed firstly turning to find the speaker's virtual embodiment and then look for the object. Similarly, the speaker cannot easily look where they are pointing and simultaneously look at other's virtual embodiments.

3.2.10 Turn Taking

According to Markel (1975), a speaking turn begins when one of the interlocutors (speakers) starts solo talking and ends when a different interlocutor starts solo talking. When one interlocutor breaks into the turn of another they produce simultaneous

speech, but it does not end the turn of the first interlocutor. Switching pauses are assigned to the preceding turn. A switching pause is the period of silence from the end of one interlocutor's solo talking to the beginning of his partner's solo talking. There are two types of simultaneous speech. Overtalk, which is defined as simultaneous speech where the first interlocutor does not give up his turn and continuous solo talking. Switching overtalk, which is defined as simultaneous talk where the first interlocutor gives up the floor and simultaneous speech ends as his partner begins solo talking. In the distribution of co-verbal behaviour individual selection can play a great role because the individual has a wider range of options for expression than with verbal behaviour. Markel holds that the most fruitful co verbal behaviours for understanding and observing social interaction are head nodding, face looking, smiling, head touching, and speaking, including simultaneous speech.

CVEs do not generally allow for head nodding, face looking, smiling, or head touching, leaving the participants with only the verbal options to initiate and indicate turn taking. Verbal turn taking seems therefore in important element in the support for collaboration in CVEs. Turn taking in CVEs has been researched by Bowers, Pycock and O'Brien (1995). The paper reports the use of qualitative, interpretative analyses of social interaction to reveal some systematic problems with turn taking and participation in CVEs. The authors found many examples during their analyses of CVE collaboration, of familiar and expected human-human collaboration behaviours. However, the authors show and conclude that managing turn taking smoothly is problematic for CVE users, and that this problem is exacerbated by poor audio connections and insecurities about the reliability of the hardware connections between users. They also found evidence for the fact that CVE users try to find ways to

collaborate within the limits of CVE technology, and that users try to adapt their behaviour accordingly to get on with their task.

3.2.11 Unavailable Collaborative Actions

Some collaborative actions are unavailable in the CVEs under research in this thesis. Examples are eye-contact and gaze, touch, facial expressions and dress. The fact that they are absent makes them important to discuss in terms of the impact that this absence may have on CVE collaboration in particular.

Gaze

According to Argyle (1975), eye contact is important to the regulation of flow of conversation, giving feedback to speaker about communication; expressing emotions, and informing participants about the nature of the relationship. Similarly, the direction in which the eyes are pointed, called gaze, is important. Gaze, the direction of the gaze, duration of the gaze and immediacy of directing a gaze, indicates where attention is aimed. In CVEs users have no direct eye contact with other users, and no real feedback about where the other users' gaze is directed. Instead they are presented with the gaze of the virtual embodiment, which may or may not be where the real focus of attention lies of the user who controls the virtual embodiment. Although this may potentially create confusion for users new to the concept of CVE collaboration, it is to be expected that users adapt to these differences.

Touch, Facial Expressions, and Dress

Touch, facial expressions and dress are not really described here because they are not represented in the CVEs under research and will consequently not really be part of the

observations. However, this does not mean that other CVEs may not include some of these features. There are technical efforts underway to introduce the sense of touch, and facial expressions into virtual environments. Other technological efforts include the rendition of a true 3D likeness of oneself for use in CVEs, and user tailorable clothing for virtual embodiments.

3.2.12 Trust Building: Interface Competence

In order to be taken serious, during interaction, people have to establish and maintain a set of social norms or etiquette of social conduct during the interaction (Nohria and Eccles, 1992). There is some evidence that in order to establish and confirm ones perceived trustworthiness as a competent computer-mediated collaborator, a participant has to be perceived by the other participants as acting according to the social norms (Rocco, 1998), which are applicable within the mediated context.

In CVEs, participants are only aware of each other's voice and virtual embodiments, and the co-verbal cues which can be observed from these. The interpretation and display of proxemic shifts, and the observation of the spatial regulation rules are influenced by the ability to fine-tune navigation and positioning of ones virtual embodiment. Conversation, turn-taking, and the display and perception of co-verbal cues are influenced by the performance and reliability of the audio-channel. And last but not least, the success of the collaboration is heavily influenced by the success of the phatic communication, which in its turn is dependent on the performance and perceived reliability of the audio-channel and the participants' ability to coordinate their embodiments. However, in a CVE, users have to manage this using a much narrower bandwidth than available in real world settings. A rich CVE interface, with

effective virtual embodiments and easy to use navigation controls should help participants to reach the working consensus for the interaction, especially in those situations where the role structure is initially not obvious to the participants.

There are several stages during the CVE interaction at which the success of the collaboration is at risk (see figure 3.2). In order for the participants to be able to acknowledge each other's presence, they have to be able to either hear or see each other. In order to be able to fine-tune their position in the virtual space, they have to be reasonably competent at navigation. In order to be able to contribute to the meta-planning of the collaborative task, they have to be relatively undistracted by concerns of audibility, visibility, and sociability. In order to collaborate on the activity at hand, they have to be certain that they are audible and visible to the other participants when they think they are and vice versa.

3.3 Mediated Collaboration

The verbal and nonverbal communication activities summarized above can serve as criteria for a rough estimation of usability of technology mediated communication systems. Table 3.2 juxtaposes the different mediated communication systems discussed in Chapter 2, with the ideal real life situation. It has to be noted that the scores are not directly based on quantitative data, but rather an estimation made by the author and derived from the available literature.

Social Behaviour	Real Life	Audio Conferencing	Video Conferencing	Media Spaces	Virtual Conferencing
Verbal communication	*****	****	***	**	**
Phatic communication	*****	****	****	****	****
Spatial regulation	*****	-	*	*	****
Proxemic shifts	*****	-	-	-	*****
Co-verbal behaviour	*****	-	***	***	*
Turn-signal	*****	**	**	**	**
Peripheral Awareness	*****	-	*	**	*****
Trust Building	*****	****	****	****	***
Reciprocity	*****	**	**	***	*****
Indexicality	*****	**	**	***	*****

Table 3.2: Potential success of social behavioural repertoire during technology mediated human-human interaction. Scores are from low fidelity (*) to high fidelity (*****), or not available (-).

Additionally, it has to be noted that the scores for Virtual Conferencing (such as takes place in CVEs) is probably optimistically high.

3.4 Conclusions

CVE participants do not perceive each other directly at all, instead they interact via their audio signals and their virtual embodiments (VBs or "avatars"). These VBs have a limited capacity for expression and do not automatically relay the real expressions that commonly accompany social interaction of their owner. The review of social behaviours has clarified that CVE users need to be supported in terms of collaboration support, which is making visible the conduct of the user in and with the CVE, to others and to themselves (see figure 3.4).

```
┌──────────────────────────────────────────────────────────────────────────┐
│  ┌─────────────────────┐    ┌─────────────────────┐    ┌──────────────────────┐ │
│→ │ Use Auditory and    │ →  │ Guide or Follow View│ →  │ Employ or Perceive   │ │
│  │ Visual Input to     │    │ to Locus of         │    │ Recipient-Designed   │ │
│  │ Monitor space       │    │ Attention           │    │ References           │ │
│  └─────────────────────┘    └─────────────────────┘    └──────────────────────┘ │
└──────────────────────────────────────────────────────────────────────────┘
```
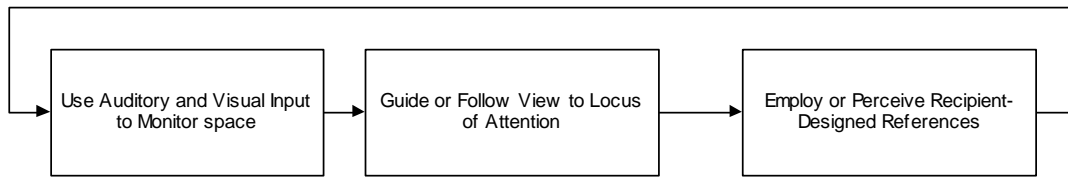
Figure 3.4: Elements of conduct of participants in collaboration.

This blow-by-blow account of the interactions of a user or users to engage in an interaction with another user can be fitted into a more general schema of collaboration (see figure 3.5).

```
┌──────────────────────────────────────────────────────────────────────────────────┐
│  ┌───────────────────────────────────────┐   ┌───────────────────────────────────────┐ │
│  │ ┌─────────────────┐  ┌─────────────────┐ │   │ ┌─────────────────┐  ┌─────────────────┐ │ │
│  │ │Use Auditory and │→ │Guide or Follow  │ │   │ │Employ or Perceive│→│Collaboration    │ │ │
│  │ │Visual Input to  │  │View to Locus of │ │   │ │Recipient-Designed│ │Proper           │ │ │
│  │ │Monitor space    │  │Attention        │ │   │ │References        │ │                 │ │ │
│  │ └─────────────────┘  └─────────────────┘ │   │ └─────────────────┘  └─────────────────┘ │ │
│  │  Peripheral Awareness  Interactional     │   │  Focused Awareness   Interactional       │ │
│  │                        Transition        │   │                      Transitions        │ │
│  │  Unfocused Collaboration                 │   │  Focused Collaboration                   │ │
│  └───────────────────────────────────────┘   └───────────────────────────────────────┘ │
│  Meta Collaboration                                                                      │
└──────────────────────────────────────────────────────────────────────────────────┘
```
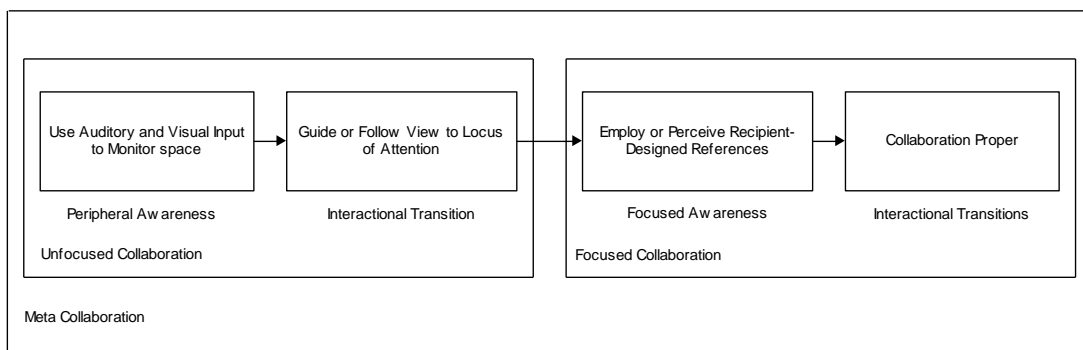
Figure 3.5: General schema of collaboration.

The human capacity for peripheral awareness and notions of reciprocity are limited by how the happenings inside the CVE are portrayed to them. If a user performs an action on an object inside the CVE, this will not be visible to the other participants in the CVE unless the act is made visible to them. The effect an action will have needs to be made visible to the acting user, but also to the other CVE participants, if they are to know that the action had an effect at all. If an action of one user has a direct effect on other users or objects, the affected individuals will not be immediately aware of this unless it is made clear to them by the CVE system, etc. There are also issues of peripheral awareness of the real space in which each CVE user is present. For

instance, local real life happenings are intrusions and distractions to the focus of the attention of the users on the happenings in the CVE, and this temporary lack of attention is not readily perceivable to the other users.