



- D3.5, Usage Evaluation of the Online Applications

Part D: Theory and Method Description for Spatial and Temporal Activity Charting of CVE Interactions

Project Number:	ACTS Project N. AC040
Project Title:	COVEN - Collaborative virtual environments
Deliverable Type:	P*

CEC Deliverable Number:	A040-UCL-CS-DS-P-035d.b1
Contractual Date of Delivery to the CEC:	31 st October 1998
Actual Date of Delivery to the CEC:	6 th November 1998
Title of Deliverable:	Usage Evaluation of the Online Applications
Work package contributing to the Deliverable:	WP3
Nature of the Deliverable:	R**
Author(s):	Jolanda Tromp (Nottingham)

Abstract: This paper describes the theoretical considerations and method for the scoring of interactional processes of small groups in CVEs. The method explains how the observations from video recordings of the networked COVEN trials can be scored along a time axis and the elements of behavior are coded using the categories defined here.

Keyword list:

Collaborative Virtual Environments, Usability Inspection, Presence, Cognitive Walkthrough, Heuristic Evaluation

***Type: P-public, R-restricted, L-limited, I-internal**

****Nature: P-Prototype, R-Report, S-Specification, T-Tool, O-Other**

1	INTRODUCTION	1
2	THEORY FOR SPATIAL AND TEMPORAL ACTIVITY CHARTING.....	2
2.1	THE OBSERVERS ROLE	2
2.2	THE INTERACTION CATEGORIES	2
2.3	THE UNIT TO BE SCORED	3
2.4	SCORING PROCEDURE.....	3
3	UNITS OF ANALYSIS FOR OBSERVATION.....	4
3.1	SPATIAL REGULATION	5
3.2	PROXEMIC SHIFTS	6
3.3	COVERBAL BEHAVIOR.....	7
3.4	PHATIC COMMUNICATION.....	7
4	THE CATEGORIES	10
4.1	COMMUNICATE (C)	10
4.2	MANIPULATE (M).....	12
4.3	NAVIGATE (N)	13
4.4	POSITION (P)	13
4.5	SCAN (S).....	14
5	CONCLUSIONS.....	16
6	REFERENCES.....	17

1 Introduction

This document describes the theoretical considerations and the method to score the temporal and spatial activities of user of Collaborative Virtual Environments (CVE). The method explains how to score the interaction of the participants which are recorded over time, using an integration of several theories of collaboration. This method is different from other observation methods in that strenuous efforts are made to clarify the bases upon these inferences are made, to cancel out the effects of value judgments from the observer's own particular point of view, to standardize the process of inference, and to determine whether the operation is reliable, by testing and refining the categories of behavior which are inquiry.

Interaction process analysis is a method described by Bales (1951) for use in the original observation of small group interaction as it occurs. The heart of the method is a way of classifying direct face-to-face interaction as it takes place, act by act, and a series of ways of summarizing, analyzing the resulting data so that they yield useful information (Bales, 1951). An observer scores the occurrence of small group behaviors as they occur, based on a system of pre-defined categories. The number of times the categories are scored and the order in which this occurs gives us information about the behavioral patterns of the member of the small group under observation. This method provides a quick way of collecting quantitative data from observations. For COVEN these observations are made from video-recordings of CVE interaction.

2 Theory for Spatial and Temporal Activity Charting

A small group is defined as any number of persons engaged in interaction with each other in a single face-to-face meeting or a series of such meetings, in which each member receives some impression or perception of each other member distinct enough so that he can, either at the time or in later questioning, give some reaction to each of the others as an individual person, even though it be only to recall that the other was present. A number of assumptions are included in this statement (paraphrasing Bales):

Assumption 1: All small groups are similar in that they involve a plurality of persons who have certain common tasks arising out of their relation to an outer situation, and certain social and emotional relationships arising out of their contact with each other.

Assumption 2: Each act of each individual in the group can be analysed with regard to its bearing on these tasks and the observable actions.

Assumption 3: In any given interaction the group member to whom the actor is talking is presumably trying to put herself in the actor's shoes, and by this process the group member is attempting to empathize with the actors and, at the same time, is testing her own reaction to what she perceives - all of this as a basic process in communication.

2.1 The Observers Role

The observer has a list of categories into which he classifies every item of behavior he can observe and interpret. The classification he makes, involves the imputation of meaning, the "reading in" of content, the inference that the behavior has function(s), either by intent or effect.

The observer is familiar with the categories in great detail. He knows the central theoretical meanings of the variables or categories and he also knows the great range of variations of concrete behavior included within each of these categories. He has practiced and developed certain sensitivities and a facility in making rapid decisions and putting down scores. As the people in the observed group talk to each other, the observer breaks their behavior down into the smallest meaningful unit until he can distinguish, and records the scores by putting down beside the proper category the number of the person speaking and the number of the person spoken to. The observer follows the interaction continuously in this microscopic manner, attempting to keep the scores in the sequence in which they occur, and to omit no item of behavior. All kinds of behavior - overt skeletal, verbal, gestural, expressive - are included, provided that the observer can assign meaning to the behavior in terms of the categories.

Reliability of unitizing depends very heavily on the training of the observers and on the development in joint training sessions of certain minor conventions which help adapt the method to the particular scoring situation. The observers should develop a theoretical understanding and a "feel" for the placement of the categories on the scoring form. The observers should be familiar with the specific content of each category and the definition of the units to be scored. This document is the first iteration in the development of such a scoring method to look into the collaborative, communicative, and interactive behaviors of multiple users involved in a task in a CVE.

2.2 The Interaction Categories

The whole theoretical rationale of the method is involved one way or another in the categories and their content and arrangement.

Bales set of categories is meant to be a general-purpose framework for observation which can be used to obtain a series of standard indices regarding the structure and dynamics of interaction in any small group. The set of categories used for scoring interaction must be concerned with aspects of interaction so general that they will appear in communication between members of any small group, regardless of the idiosyncratic content of the topic of their discussion or the kind of concrete problems or subjects with which they may be dealing. In addition to the formulation of behavior that always appears, the list must be concerned with certain variations of behavior which may not be frequent in certain groups but which potentially can and do appear under certain conditions, regardless of idiosyncratic content.

All of the categories included should assume essentially the same time span; that is, they should all refer to single acts of communication or expression.

The total set of categories used for observation expresses a conception of the various elements in interaction systems as we observe them on a relatively low level of abstraction. The categories fit together so that, even without theoretical explanation, they can be grasped and held in mind as a gestalt or “total map”. Each category is meant to gain its central meaning from its position in the set of categories. The placing of a category in a particular position with regard to the other categories is the most important part of its definition.

2.3 The Unit to be Scored

The unit to be scored is the smallest discriminable segment of verbal or nonverbal behavior to which the observer, using the set of categories after appropriate training, can assign a classification under conditions of continuous serial scoring. This unit may be called an act, or more properly, a single interaction, since all acts in the present scheme are regarded as interactions. The unit as defined here has also been called the single item of thought or the single item of behavior.

2.4 Scoring Procedure

First, all the actors in the observation get a number. Secondly is scored who acts in reaction to whom. Whenever the act seems to be addressed to no particular other, or to more than one other, the symbol “0” is used to designate the target. In cases where more than two persons act at once, as when they all laugh together, or shout together, the symbol “0”) is used to designate the actor. Thus a general laugh would be scored 0-0 (all to all). A general nodding of heads to actor 1 would be scored 0-1 (all to one). If the actor is giving his attention primarily to the situation - the outer situation - the act is recorded 1-x, even though he may be speaking aloud so that the other members can hear. This then, is noted down with (category 1, 1-0) or in the appropriate category for the observed behavior (scores for who to whom along a list of the categories).

Bales' system assumes that a given temporal segment of behavior will be classified in only one category. The frequency with which classification dilemmas arise is in part a function of the fundamental soundness of the underlying rationale from which the dimensions are derived. Bales notes that the question of reconciling interpretation dilemmas is a question of: (1) what does the manual say? and (2) what modifications of the manual are required to anticipate issues of this nature in the future?

3 Units of Analysis for Observation of Collaborative Behavior in CVE's.

Duncan (1975), has defined units for analysing social interaction in one-to-one, face-to-face conversations which segment speaking turns. The following behaviors were found to accompany a change in speaking-turn: a) an unfilled pause; b) the turning of the speaker's head towards the auditor; c) a drop in paralinguistic pitch and/or loudness in conjunction with a phonemic clause, either across the entire clause, or across its final syllable or syllables; d) a relaxation of the foot or feet of the speaker from a marked dorsal flexion; e) an audible inhalation; f) the use of any pitch level-terminal juncture combination other than 2 2| at the end of a phonemic clause; g) a paralinguistic drawl on the final syllable or on the stressed syllable of a clause; h) the termination of any hand gesticulation used by the speaker, or the relaxation of a tensed hand position (e.g. a fist) by the speaker; i) the use by the speaker of one of a set of stereotyped expressions, such as "but uh", "or something", or "you know", j) a drop in paralinguistic pitch and/or intensity, in conjunction with a sociocentric sequence; and k) the completion of a grammatical clause, involving a subject-predicate combination.

Using these observable behavioral units he found higher level units of interaction which segment speaking turns in one-to-one, face-to-face conversations, which he proposes are the main elements of a "speaking-turn system". The speaking-turn interaction units are: Turn-System States, Speaker-Turn Signal, Speaker-Gesticulation Signal, and Speaker-State Signal. The speaker-turn signal, speaker-gesticulation signal and speaker-state signal assist in negotiating changes in the turn-system states between the speaker and the auditor. A speaker is defined as a particular participant who claims the speaking turn at any given moment. An auditor is a participant who does not claim the speaking turn at any given moment.

The auditor may claim the turn when the speaker displays a turn-signal. The turn-signal is composed of a set of six behavioral cues, found variously in intonation, context, syntax, paralinguistic, and body-motion. The cues are behaviors f, g, h, i, j, and k. The Display of any single cue was found to be sufficient to constitute a display of the signal. The probability of an auditor turn-claiming response to a signal display was found to be a linear function of the number of cues displayed, without regard to the specific cues composing the display. However, the speaker-gesticulation signal: the hands being engaged in a gesticulation, as opposed to being engaged in a self-adaptor, or being at rest, or behavior h, was found to negate any turn signal concurrently being displayed. The speaker-state signal is hypothesized to mark a shift from auditor to speaker state and is defined for at least one of a set of two discrete cues: i) a shift in head directions, away from one pointing directly towards the vis-a-vis, and ii) initiation of a gesticulation.

A smooth speaking turn exchange involves the following ordered sequence of three events: A) the speaker displays a turn signal (and does not conjointly display a gesticulation signal); B) the auditor switches to the speaker state, displaying a speaker-state signal; and C) the previous speaker switches to the auditor state, relinquishing the turn. During speaker-auditor interaction; e.g. within in a turn, interaction units further segment the speaking turns. These within-turn units are: Auditor Back-Channel Signal, Speaker Within-Turn Signal, and Speaker Continuation Signal. The most obvious behaviors observed during speaking turns are head nods and vocalizations termed "back-channel" behaviors. Auditor back-channels are: A) verbalized signals (such as M-hm), B) Sentence completions; C) Requests for clarification; D) Brief restatement; E) head nods and shakes. Back-channel behaviors do not constitute a speaking-turn, nor are they claimings of a speaking-turn, but they were found related to a set of two cues, comprising the speaker within-turn signal. Two speaker cues result in a display of a subset of auditor back-channels: a) completion of a grammatical sentence, b) turning of the speaker's head towards the auditor. Either one of these cues is sufficient to constitute the signal. Under certain circumstances this Speaker Within-Turn Signal is followed by a Speaker Continuation Signal.

The speaker continuation signal is related to the preceding display of both a) the speaker within-turn signal, and b) the auditor back-channel signal. The speaker continuation signal is composed of one of the cues of the speaker-state signal: shift away in head direction. When no auditor back-channel intervenes between the end of one unit of analysis and the beginning of the next, the speaker-state signal is very often followed by one or both of the speaker within-turn cues. When auditor back-channels occur before the end of a unit of analysis, the probability of a subsequent speaker continuation signal significantly increased. When the back-channel occurs between units, there is no accompanying increase.

The within-turn interaction units parallel speaking-turn interaction units in their structural characteristics: a) both require the appropriate, coordinated action of both participants, and b) both involve ordered sequences of action. The beginning of a within-turn interaction unit appears to be marked by a speaker continuation

signal, analogous to the marking of a turn beginning by a speaker-state signal. The beginning of a new within-turn unit appears to be associated primarily with either one of two preceding events: a) a speaker within-turn signal, regardless of whether or not a between-unit auditor back-channel intervened, or b) an early auditor back channel, regardless of whether or not a speaker within-turn signal preceded it. The within-turn interaction unit appears to provide the participants with means by which to pace speaking turn at a rate that takes both speaker and auditor into account.

Bales' system (1951), is developed to analyse small group behaviors relevant for activities as can be observed in meetings, discussion groups, and therapy groups. The categories he has defined are: A) Social-Emotional Area: Positive; 1) Shows solidarity; 2) Shows tension release; 3) Agrees; B) Task Area: Neutral; 4) Gives suggestion; 5) Gives opinion; 6) Gives orientation; C) Task Area: Neutral; 7) Asks for orientation; 8) Asks for opinion; 9) Asks for suggestion; D) Social-Emotional Area: Negative; 10) Disagrees; 11) Shows tension; 12) A shows antagonism. Where 6 and 7 are Problems of Communication, 5 and 8 are Problems of Evaluation, 4 and 9 are Problems of Control, 3 and 10 are Problems of Decision, 2 and 11 are Problems of Tension Reduction, and finally, 1 and 12 are Problems of reintegration. A can be defined as Positive Reactions, B as Attempted Answers, C as Questions, and D as Negative Reactions.

Neale (1997) used a multiple activity chart scoring method to observe interaction with a virtual environment. She applied the scoring of observations behaviors in categories over time to several Virtual Learning environments for children with learning disabilities to six case studies. The scoring categories were derived from the Constructivist theory of learning. She coded language and actions of two simultaneous users of the software; a teacher and the child with learning disabilities.

Not all seven principles of the theory of learning are visible behaviors, so only observations of behaviors gleaned from the interactions between the teacher and the child and the VE could be scored and Neale dropped two of the seven principles. The remaining five categories were extended based on pilot studies, including subcategories within the main categories, and codes for observations of negative occurrences of the principles. For instance, the according to the Constructivist theory of learning, principle 2 of the theory states that the learning process should "2: facilitate knowledge construction, not reproduction". Within this category 5 subcategories are defined for coding, for instance "2.1: Child moves/explores", and "2.2: Child responds to constraints", etc. Also, negative occurrences are provided with a code, for instance "2.B: No response to /ignore constraints". Additional instances of observable behaviors are also coded, such as the child's frustration, and changes in attention.

With regards to the applicability of the method Neale reports that she does not know of previous investigations using this method applied to VEs. With regards to using the method she reports that some behaviors were difficult and very subjective to score, some categories were not specific enough, and some categories were extremely difficult to infer. On the whole the method assisted in uncovering a number of theoretical relationships between certain aspects of the design of the VEs and their subsequent use.

Kanuritch et.al. (1997), used a multiple activity scoring chart to score teleoperator behaviors with a remote handling system for nuclear power station maintenance and repair activities, as part of a usability study of two different handling systems. Performance data such as time to complete tasks, errors made, subjects' movement of joystick, display viewed, speech taking place and activity of robot manipulator were charted. No record exists of analysis of the multiple activities charted, nor are any remarks on the usefulness of the method available.

3.1 Spatial Regulation

According to Sheflen (1975), micro-territories can be the basis for comprehending and measuring activities and relations of activities. Territory can be defined as an area of open space which 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, but they have a wide range of duration. Spatially, these mini-territories are organized 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.

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.

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.

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

The location is that increment of space which 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 subterritories. 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 concretized 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.

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 allomorphic or 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. 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.

Vine (1975) reviewed human territoriality compared to other species' territoriality. Humans often show strong attachments to areas they use frequently, and we may monopolize 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 man even when reinforced by legitimated ownership of space or motivation to maintain privacy is typically highly ritualized 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 apologize 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 which "depersonalizes" 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 a necessity to react and interact in one way or another.

3.2 Proxemic Shifts

Erickson (1975) studied patterns of interpersonal distance in face-to-face encounters, referred to as proxemic shifts in face-to-face interaction. 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. Erickson devised a code for gross features of

proxemic shifts and a set of codes for gross features of other aspects of verbal and nonverbal behavior. Information from these codes, when combined together, provided a synoptic view of interactional structure. The behaviors coded are: speech behaviors, non-verbal behaviors, and hierarchical ordering of natural units (segments) of speaking/listening, each of these categories are divided into subgroups. 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.

3.3 Coverbal Behavior

Coverbal behavior has been researched by Markel (1975), and is defined as the behavior of interlocutors (speaker) which occurs in association with or accompanying words, but which is not essential for the articulation or grammatical functioning of those words, basically non-verbal communication accompanying verbal communication. Alternation of action and inaction, and the subsequent rotation of performance among individuals in a group is the most salient feature of group dynamics. For this reason, the conversational turn has been used as the unit of analysis of conversational interaction and group behaviors. Markel uses the following definitions: a speaking turn begins when one of the interlocutors 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 continues 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 coverbal behavior individual selection can play a great role because the individual has a wider range of options for expression than with verbal behavior. Markel holds that the most fruitful coverbal behaviors for understanding and observing social interaction are head nodding, face looking, smiling, head touching, and speaking, including simultaneous speech.

3.4 Phatic Communication

Phatic communication serves to establish and maintain a feeling of social solidarity and well-being, by exchanging words. The pervasive attitude in face-to-face interaction is that speech is only one among the many strands of communication and that the communicative function of any 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 we can understand the particular texture of interaction only by seeing the relationship of the different strands.

Three semiotic notions are used, adopted from C. S. Pierce, the symbol, the index, and the sign. The concept of the symbol is used in the common linguistic usage of linguistic elements (of varying size) being used as symbols of their referents. The index is a sign which reveals personal characteristics of the speaker. The icon is one event acting as the sign of another through similarity of structure between the two events. The most important of these concepts for the study of phatic communication is the index, because the prime function of phatic communication is the communication of indexical facts about speakers' identities, attributes, and attitudes, and that these indexical facts constrain the nature of the particular interaction.

The temporal structure of interactions is divided into three major phases, the opening phase, the medial phase, and the closing phase. The function of the behavioral activity that characterizes 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 medial phase. The closing phase is once again a transitional phase, easing the transition from full interaction to departure. Phatic communication strongly characterizes these marginal phases of communication. 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 (1959) has called the "working consensus" of the interaction, about some aspects of their working roles in those situations where the role structure is not previously obvious to the participants.

The function of phatic communication during the opening phase of interaction are the following: to provide the participants with acceptable means of starting 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 allows the participants 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. It also serves to consolidate the relationship between the two participants, by means of behavior which emphasizes 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:

- 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. This is one area in face-to-face interaction that has been studied the most.
- 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..."

When one participant is static in space, and the other is moving towards him, in whatever type of physical locale, then, unless there are overriding special reasons, there seems to be a strong tendency, both in Britain and America, for the "incomer" to initiate the exchange of phatic communication. By breaking the silence first, the speaker defines some aspects of the role he is prepared to play in the oncoming interaction.

The closing phase of interaction

- 1) The initiator of the closing phase performs appropriate signals of transition, indicating his desire for the onset of the closing phase and the end of the medial business 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 medial 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 medial phase makes no call for such an indication of apparent cordial, attentive agreement. Also, linguistic signals of transition from the medial phase include the same sort of vague, curtailed utterances as are used in the opening phase, such as "Well..."
- 2) Display of behaviors which emphasize 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 vigor of facial expressions, loudness, such as would be necessary for audibility and visibility at a greater distance.
- 3) Exchange of tokens of phatic communication, whose functional characteristics

- 4) Exchange of conventional contact gestures of parting and adoption of conventional facial expressions, where and as appropriate to the relationship. (Fourth and fifth stages are usually simultaneous.)
- 5) Exchange of conventional contact gestures of parting and adoption of conventional facial expressions, 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 reentry to earlier stages of the interaction is necessary. Conventional parting is thus a cooperative achievement obeying rather strict constraints.

Skill in managing the behavioral 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 behavior are a very important social and psychological instrument, 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 can be usefully described as a ceremony functioning as a rite de passage, easing and signaling the transitions to and from conversational interactions. The transactions during arrival and departure are called telectic rites (from the Greek concept of putting off the old and putting on the new), this seems to be an apt term for the 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.

4 The Categories

The following categories have been created based on a combination of the information above, and an observation of video footage of two participants collaborating on a virtual moving task in MASSIVE-2. Each category falls apart in activities which the actor is intending to perform. For instance, an actor may navigate (N) towards a certain position in order to be able to communicate (N-C), or an actor may communicate (C) in order to verify (V) that the other actor has perceived his/her action (C-V). For each activity, not only the fact that this act can be observed is interesting information, but also when the behavior is absent, and thus not available for observation is useful information. However, it is difficult to score the absence of behaviors, although in some instances it is possible to observe results of the absence of an act, for instance when an actor is temporarily suspending their activities in the CVE this is not immediately noticeable to the other actor(s), only when the other actor(s) try to communicate with the 'dormant' VB do they notice, and communications like "Are you there?" can be observed. The failure to notify the others of a temporary suspension of attention could be scored as a negative instance of C-V.

4.1 Communicate (C)

Actors communicate for many reasons. Communication is an essential element in collaboration and many aspects of collaboration can be found with repetition in the communications observed.

4.1.1 Communicate-Explore (C-E)

Actors communicate in order to explore the environment. They may be observed discussing areas or directions before actually taking off.

4.1.2 Communicate-Manipulate (C-M)

Actors communicate in order to manipulate an object. They may be observed discussing an object in their shared space in order to find out who is, or who will manipulate the object (first).

C-M describe location for object

C-M initiate object movement

4.1.3 Communicate-Position (C-P)

Actors communicate in order to position their VB in relation to the position of something else in the shared space. They may be observed discussing the precise alignment of their VB or that of the other actor(s).

C-P describe objects in order to make them shared objects

C-P describe objects in relation to other objects

C-P describe objects in view

C-P describe position with regards to others position

C-P describe possible object position in relation to own body

C-P describe view

C-P point at location

C-P point at object

C-P propose move of object

C-P request clarification of position intended

4.1.4 Communicate-Scan (C-S)

Actors communicate in order to scan the environment. They may be observed discussing the relative position of objects or other VBs in relation to other objects in the shared space or in relation to their own VB or objects near them.

4.1.5 Communicate-Test (C-T)

Actors communicate in order to test their communication means. They may be observed to say: "Hello, can you hear me?", or similar test messages. making sure the equipment works is an important part of the initiation of collaboration activities.

- C-T "hello" (check for presence)
- C-T "hello" testing audio adjustments
- C-T call halt to convo
- C-T check presence of other (Are you still there?)
- C-T describe physical screen (external)
- C-T request confirmation of being audible (can you hear me?)
- C-T trying out sound with jokes
- C-T use name to get attention

4.1.6 Communicate-Verify (C-V)

Actors communicate in order to verify what they have heard, seen, done or said. Making sure the other actors have perceived the same thing is an essential element of collaboration. Many different kinds of verification behaviors can be observed. It is recommended to make a record of the kinds of verification that takes place.

- C-V "hello" (acknowledge presence)
- C-V accompany virtual movement/action with noise (made by self through mike)
- C-V agree
- C-V agree on seeing the same thing (non-shared object = physical screen item)
- C-V agree on view
- C-V agree to start
- C-V announce disengagement with CVE
- C-V announce end of position action
- C-V announce lostness
- C-V announce mistake
- C-V announcement of absence (afterwards)
- C-V ask for clarification (more details)
- C-V call each other
- C-V catch up (what did you want me to look at?)
- C-V check problems/progress (oh what happened?)
- C-V confirm intentions
- C-V confirm utterance
- C-V describe directions
- C-V describe intentions with shared objects
- C-V devide the work
- C-V discuss shared object
- C-V discuss until agreed
- C-V fail to announce lostness (negative)

C-V greetings
C-V monitor actions by speaking outloud when split up in subgroups
C-V question activity
C-V request clarification of intentions
C-V request clarification of meaning
C-V request clarification of utterance
C-V request confirmation of happening (oh are you doing that?)
C-V request confirmation of utterance
C-V shout of surprise at seeing each other
C-V surprise at absence (didn't notice)
C-V talk about mike levels (bad)

4.2 Manipulate (M)

Actors manipulate objects in their environment for a variety of reasons. Depending on the object being manipulated the goal of their manipulations may be more or less clearly determined. Objects may also simply be manipulated in the space in order to move them from A to B, or in order to move them out of the view or path of the actor, or in order to test the effects and mechanisms of the manipulation act.

4.2.1 Manipulate-Communicate (M-C)

Objects may be manipulated in order to communicate something about them, such as the position, the person who controls the manipulation, to signify the particular object, or to draw attention to the object.

M-C manipulate object between each other

4.2.2 Manipulate-Navigate (M-N)

Objects may be manipulated in order to navigate them through the environment. The actor picks up the object and carries it from A to B.

4.2.3 Manipulate-Operate (M-O)

Objects may be manipulated in order to operate the available functions on them.

4.2.4 Manipulate-Position (M-P)

An object may be manipulated in order to position it carefully in an intended position or in order to get it out of its original position.

M-P move object

M-P position object

M-P try to move object

4.2.5 Manipulate-Scan (M-S)

Objects may be manipulated in order to scan them or the space the object was obstructing.

4.2.6 Manipulate-Test (M-T)

Objects may be manipulated in order to test the actors manipulation skills and abilities or in order to test the particular manipulation actions which can or must be performed on the object in order to operate it.

M-T try out body movement functions on other

4.3 Navigate (N)

Actors navigate for many reasons through the CVE. Navigation is defined as directed movement from one spot to another, with a clear start and finish. Navigation can be differentiated from 'Scanning' behavior by its movement from A to B, where scanning involves looking from A to B, without obvious relocation of the VB.

4.3.1 Navigate-Communicate (N-C)

Actors navigate in order to communicate. They may specifically try to find other actors, and position their VB to initiate communication.

N-C approach object with other in view (keeping it there)

N-C approach other

N-C approach other

N-C approach other with object in view (keeping it there)

4.3.2 Navigate-Explore (N-E)

Actors navigate in order to explore the environment. They may be seen moving from A to B and back again, to gain an understanding of the space and their position in it in relation to other actors and objects.

4.3.3 Navigate-Find (N-F)

Actors navigate in order to find a certain object. They should know approximately where the object is, but they still have to navigate their VB around the space in order to locate it.

N-F find other

N-F move view to encompass object

4.3.4 Navigate-Manipulate (N-M)

Actors navigate in order to manipulate an object in the environment. This could be a door which the actor needs to go through, or an object on which the actor wishes to act.

N-M approach object

N-M approach object

4.3.5 Navigate-Position (N-P)

Actors navigate in order to position either themselves or an object in their possession. They position their VB in order to act onto something, and the positioning may have to be performed very precise.

N-P fall out of room (both) (negative)

N-P get other and object in view

N-P split up in subgroups

4.3.6 Navigate-Scan (N-S)

Actors navigate the environment in order to scan from A to B. In order to be able to scan effectively they may have to navigate to a comfortable position first, but without actually staying in this position very precisely. They may be seen navigating to locate A and B into one view, or test a sweep from A to B, or back up to the furthest corner for the largest overview of the space.

N-S back up

N-S back up for "big picture"

4.4 Position (P)

Actors position themselves or objects in the environment for a variety of reasons. Positioning is defined as the precise act of alignment of the VB or the object in relation to something else in the space.

4.4.1 Position-Communicate (P-C)

Actors may position their VB in order to communicate with the other actor(s). They can usually be observed positioning themselves facing the other actor(s) they are communicating with, but they may also try to face the object they are communicating about.

P-C face each other

P-C move back view to encompass embodiment of other

P-C other lost from view (negative)

4.4.2 Position-Explore (P-E)

Actors may position themselves in order to explore the environment. They may be observed trying to encompass as much of the environment in their view as possible in order to explore the relationship between objects in the space.

4.4.3 Position-Manipulate (P-M)

Actors may position themselves in relation to an object in order to reach the most advantageous position to manipulate the object.

4.4.4 Position-Scan (P-S)

Actors may position themselves in order to scan the environment. They may be observed positioning their VB in order to scan from A to B and back.

P-S approach other and object

4.4.5 Position-Verify (P-V)

Actors may position themselves in order to verify their position in relation to others or their own location in relation to objects and the space to themselves. They may be observed to use certain objects to fix their position.

P-V approach object while keeping other in view

P-V request directions

4.5 Scan (S)

Actors scan the environment for many different reasons. Scanning is defined by a sweeping movement from A to B, not with the intention of moving the whole VB but with the intention of viewing the (part of) space. Actors may be observed trying to sweep from A to B several times before they can actually perform a smooth sweep.

4.5.1 Scan-Communicate (S-C)

Actors scan the environment in order to communicate a location or intention in relation to the location of something else. They may be observed using their viewpoint and a sweep from A to B as an aid in the referencing of locations A and B even if the other actor(s) are not capable of seeing this scanning motion.

S-C keep other in view for very few seconds

4.5.2 Scan-Find (S-F)

Actors scan the environment in order to find something or somebody. They may be seen trying to understand their position in relation to other objects in the space, or trying to locate something in relation to themselves or something else.

S-F find next object

4.5.3 Scan-Manipulate (S-M)

Actors scan the environment in order to manipulate an object. They may be observed trying to scan the space in order to determine to see what they need to see to manipulate the object successfully.

4.5.4 Scan-Navigate (S-N)

Actors scan the environment in order to navigate to a certain location. They may be observed trying to scan the environment in order to determine the shortest path to the desired goal.

4.5.5 Scan-Position (S-P)

Actors scan the environment in order to position themselves in relation to an object or other actor in the space, in order to determine the location of themselves or an object in the space in relation to themselves.

4.5.6 Scan-Sweep A-B, B-A, (S-S)

Actors scan the environment in order to make a clean sweep from A to B, and back again from B to A (where A or B may be an object, an other actor or a location). They may be observed making this sweeping motion several times during a discussion of A or B. Performing a clean sweep takes an intuitive knowledge of the distance between A and B and the actors can be seen to practice the sweep several times before getting it down to a perfect motion. These sweeps can be full-circle, 180 degrees, a quarter of a circle, and 3/4s.

S-S quick scan quick zig-zag

S-S scan 3/4, 2/4, 1/4

S-S scan from object to intended position (sweep)

S-S scan from object under scrutiny to other (sweep)

S-S scan from other to intended position

S-S scan from other to object under scrutiny (sweep)

S-S scan full circle

S-S sweep back to get other in view again (2 secs)

S-S sweep back to object

S-S sweep back to object of discussion

S-S sweep back to other

S-S sweeping back-up to furthest corner position (largest overview)

5 Conclusions

The categories can be structured and displayed in the following comprehensive way:

	Communicate	Manipulate	Navigate	Position	Scan	Verify
Communicate		C-M	C-N	C-P	C-S	C-V
Manipulate	M-C		M-N	M-P	M-S	M-V
Navigate	N-C	N-M		N-P	N-S	N-V
Position	P-C	P-M	P-N		P-S	P-V
Scan	S-C	S-M	S-N	S-P		S-V
Verify	V-C	V-M	V-N	V-P	V-S	

Work in this area is still exploratory, and more controlled experimental approaches are needed to confirm or refute the representativeness of the created categories. The categories will be pilot tested on observed data and the validity will be established. At present any conclusions as to the effectiveness of this method are inevitably mostly speculative, particularly so because appeal has been made to the conscious intuitive interpretations of participant observers to derive the categories in the first place. This text is thus intended to encourage exploratory inquiry, rather than to offer definitive conclusions.

6 References

- Bales, R.F., (1951). *Interaction Process Analysis; A Method for the Study of Small Groups*, Addison-Wesley Press, Cambridge.
- Duncan, S. Jr., (1975). Interaction Units during Speaking Turns in Dyadic, Face-to-Face Conversations, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), *Organization of Behavior in Face-to-Face Interaction*, Mouton Publishers, Den Hague, pp. 199-213.
- Erickson, F., (1975). One Function of Proxemic Shifts in Face-to-Face Interaction, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), *Organization of Behavior in Face-to-Face Interaction*, Mouton Publishers, Den Hague, pp. 175-187.
- Kanuritch, N., Farrow, G, Pegman, G.J., Prof Wilson, J.R., Dr. Cobb, S., Crosier, J, Webb, P., (1997). *A Study of the Safety and Usability of Teleoperated Remote Handling Systems*, final report UKRL.97.023.
- Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), (1975). *Organization of Behavior in Face-to-Face Interaction*, Mouton Publishers, Den Hague.
- Markel, N.N., (1975). Coverbal Behavior Associated with Conversational Turns, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), *Organization of Behavior in Face-to-Face Interaction*, Mouton Publishers, Den Hague, pp. 189-197.
- Neale, (1997). *A Structured Evaluation of Virtual Environments in Special Needs Education*, BEng Honours degree dissertation in Manufacturing Engineering and Management, University of Nottingham.
- Poyatos, F. (1975). Cross-Cultural Study of Paralinguistic “Alternants” in Face-to-Face Interaction, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), *Organization of Behavior in Face-to-Face Interaction*, Mouton Publishers, Den Hague, pp. 285-314.
- Schefflen, A.E., (1975). Micro-territories in Human Interaction, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), *Organization of Behavior in Face-to-Face Interaction*, Mouton Publishers, Den Hague, pp. 159-173.
- Vine, I., (1975). Territoriality and the Spatial Regulation of Interaction, in: Kendon, A., Harris, R.M., Ritchie Key, M., (eds.), *Organization of Behavior in Face-to-Face Interaction*, Mouton Publishers, Den Hague, pp. 357-387.

Appendix A: Glossary

Paralanguage: optional vocal effects (as tone of voice) that accompany or modify the phonemes of an utterance and that may communicate meaning.

Phatic: revealing or sharing feelings or establishing an atmosphere of sociability rather than communicating ideas.

Proxemics: the study of the nature, degree, and effect of the spatial separation individuals naturally maintain (as in various social and interpersonal situations) and of how this separation relates to environmental and cultural factors.

Appendix B: List of temporal and spatial actions observed during Virtual Moving Company Experiment with 2 participants.

“hello” (acknowledge presence)

“hello” (check for presence)

“hello” testing audio adjustments

accompany virtual movement/action with noise (made by self through mike)

adjust sound settings

agree

agree on seeing the same thing (non-shared object = physical screen item)

agree on view

agree to start

announce disengagement with CVE

announce end of position action

announce lostness

announce mistake

announcement of absence (afterwards)

approach object

approach object

approach object while keeping other in view

approach object with other in view (keeping it there)

approach other

approach other

approach other and object

approach other with object in view (keeping it there)

ask for clarification (more details)

back up

back up for “big picture”

call each other

call halt to convo

catch up (what did you want me to look at?)
 check presence of other (Are you still there?)
 check problems/progress (oh what happened?)
 confirm intentions
 confirm utterance
 describe directions
 describe intentions with shared objects
 describe location for object
 describe objects in order to make them shared objects
 describe objects in relation to other objects
 describe objects in view
 describe physical screen
 describe position with regards to others position
 describe possible object position in relation to own body
 describe view
 devide the work
 discuss shared object
 discuss until agreed
 face each other
 fail to announce lostness
 fall out of room (both)
 find each other
 find next object
 find other
 get other and object in view
 greetings
 initiate object movement
 keep other in view for very few seconds
 manipulate object between each other
 monitor actions by speaking outloud when split up in subgroups
 move back view to encompass embodiment of other
 move object
 move view to encompass object
 navigate
 other lost from view
 point at location
 point at object
 position object
 propose move of object
 qucik scan quick zig-zag

question activity
 request clarification of position intended
 request clarification of intentions
 request clarification of meaning
 request clarification of utterance
 request confirmation of being audible (can you hear me?)
 request confirmation of happening (oh are you doing that?)
 request confirmation of utterance
 request directions
 scan
 scan 3/4, 2/4, 1/4
 scan from object to intended position (sweep)
 scan from object under scrutiny to other (sweep)
 scan from other to intended position
 scan from other to object under scrutiny (sweep)
 scan full circle
 shout of surprise at seeing each other
 sound adjustments
 split up in subgroups
 surprise at absence (didn't notice)
 sweep back to get other in view again (2 secs)
 sweep back to object
 sweep back to object of discussion
 sweep back to other
 sweeping back-up to furthest corner position (largest overview)
 talk about mike levels (bad)
 try out body movement functions on other
 try to move object
 trying out sound with jokes
 use name to get attention

Appendix C: List of Observed Actions Fitted into Proposed Categories

C-M describe location for object
 C-M initiate object movement
 C-P describe objects in order to make them shared objects
 C-P describe objects in relation to other objects
 C-P describe objects in view
 C-P describe position with regards to others position

C-P describe possible object position in relation to own body
 C-P describe view
 C-P point at location
 C-P point at object
 C-P propose move of object
 C-P request clarification of position intended
 C-T "hello" (check for presence)
 C-T "hello" testing audio adjustments
 C-T call halt to convo
 C-T check presence of other (Are you still there?)
 C-T describe physical screen (external)
 C-T request confirmation of being audible (can you hear me?)
 C-T trying out sound with jokes
 C-T use name to get attention
 C-V "hello" (acknowledge presence)
 C-V accompany virtual movement/action with noise (made by self through mike)
 C-V agree
 C-V agree on seeing the same thing (non-shared object = physical screen item)
 C-V agree on view
 C-V agree to start
 C-V announce disengagement with CVE
 C-V announce end of position action
 C-V announce lostness
 C-V announce mistake
 C-V announcement of absence (afterwards)
 C-V ask for clarification (more details)
 C-V call each other
 C-V catch up (what did you want me to look at?)
 C-V check problems/progress (oh what happened?)
 C-V confirm intentions
 C-V confirm utterance
 C-V describe directions
 C-V describe intentions with shared objects
 C-V devide the work
 C-V discuss shared object
 C-V discuss until agreed
 C-V fail to announce lostness (negative)
 C-V greetings
 C-V monitor actions by speaking outloud when split up in subgroups
 C-V question activity

C-V request clarification of intentions
 C-V request clarification of meaning
 C-V request clarification of utterance
 C-V request confirmation of happening (oh are you doing that?)
 C-V request confirmation of utterance
 C-V shout of surprise at seeing each other
 C-V surprise at absence (didn't notice)
 C-V talk about mike levels (bad)
 E adjust sound settings (external)
 E sound adjustments (external)
 M-C manipulate object between each other
 M-P move object
 M-P position object
 M-P try to move object
 M-T try out body movement functions on other
 N navigate
 N-C approach object with other in view (keeping it there)
 N-C approach other
 N-C approach other
 N-C approach other with object in view (keeping it there)
 N-F find each other
 N-F find other
 N-F move view to encompass object
 N-M approach object
 N-M approach object
 N-P fall out of room (both) (negative)
 N-P get other and object in view
 N-P split up in subgroups
 N-S back up
 N-S back up for "big picture"
 P-C face each other
 P-C move back view to encompass embodiment of other
 P-C other lost from view (negative)
 P-S approach other and object
 P-V approach object while keeping other in view
 P-V request directions
 S scan
 S-C keep other in view for very few seconds
 S-F find next object
 S-S quick scan quick zig-zag

S-S scan 3/4, 2/4, 1/4
S-S scan from object to intended position (sweep)
S-S scan from object under scrutiny to other (sweep)
S-S scan from other to intended position
S-S scan from other to object under scrutiny (sweep)
S-S scan full circle
S-S sweep back to get other in view again (2 secs)
S-S sweep back to object
S-S sweep back to object of discussion
S-S sweep back to other
S-S sweeping back-up to furthest corner position (largest overview)