Modeling an Agent for Intelligent Tutoring in 3D CSCL based on Nonverbal Communication

Adriana Peña Pérez Negrón[1], Raúl A. Aguilar Vera[2], and Elsa Estrada Guzmán[1]

[1] CUCEI - Universidad de Guadalajara,
Blvd. Marcelino García Barragán #1421,
44430 Guadalajara, Mexico
adriana@sicenet.com.mx; elsa.estrada@red.cucei.udg.mx
[2] Mathematics School - Universidad Autónoma de Yucatán,
Periférico Norte Tablaje 13615, 97110, Mérida, Mexico
avera@tunku.uady.mx

Abstract. During collaboration, people's nonverbal involvement mainly intention is the achievement of the task at hand. While in 3D Collaborative Virtual Environments (CVE) the users' graphical representation −their avatars, are usually able to display some nonverbal communication (NVC) like gazing or pointing, in such a way that their NVC cues could be the means to understand their collaborative interaction; its automatic interpretation in turn may provide a virtual tutor with the tools to support collaboration within a learning scenario. In order to model a virtual tutor for 3D collaborative learning environments, based on literature review, the NVC cues to be collected; how to relate them to indicators of collaborative learning as participation or involvement; and to task stages (i.e. planning, implementing an evaluating) are here discussed. On this context, results from collecting NVC cues in an experimental application during the accomplishment of a task are then analyzed.

Keywords: CSCL, collaborative virtual environments, collaborative interaction, nonverbal communication, intelligent tutoring.

1 Introduction

There seems to be a general agreement on the motivational impact of virtual reality (VR) in the students, but there are other important reasons to use it for learning purposes. VR is a powerful context, in which time, scale, and physics can be controlled; where participants can have entirely new capabilities, such as the ability to have any object as a virtual body or to observe the environment from different perspectives; in virtual environments (VE) materials do not break, are dangerous or wear out. Also, VR allows safe experiences of distant or dangerous locations and processes [2, 18, 20].

Socio-constructivism is the fundamental theory that motivates educational uses for Collaborative Virtual Environments (CVE) [4]. Because group work improves cognitive development as well as social and management skills, CVE have the potentials to enable innovative and effective education, involving debate, simulation, role-play, discussion groups, problem solving and decision-making in a group content.

In the joint effort to solve a problem or to take care of a task, the students will interact with each other; interaction takes place when an action or its effects are perceived by at least one member of the group other than the one who carried out the action [12]. The analysis of collaboration thus can be conducted through the observation of the interaction that affects the collaborative process. However, its automatic analysis is not trivial; a main challenge consists on computationally understanding and assessing it [11]. By addition, even that computers can record every student intervention, the completely understanding of unstructured dialogue has not being accomplished in Computer Supported Collaborative Learning (CSCL) yet [17].

A number of approaches have been proposed to monitor collaboration [17] mainly applicable to conventional interfaces less naturals than expected for a CVE and not easy to adapt. For example, with menus that in a VE will cover part of the view and may be difficult to operate considering that the user has to operate his graphical representation and very probably some objects too. Or based on text communication analysis, while oral communication is substituting text in VR applications, plus VEs allow other communication channels. As a result, these approaches may not appropriately fit CVEs.

CVEs bring remote people and remote objects together into a spatial and social proximity [21], providing a technology that supports interaction through auditory and visual allowing. In such a way that, this VEs visual characteristic, guided us to explore the users' avatar nonverbal communication (NVC). There are three different approaches to transmit NVC to a VE:

1) Directly controlled –with sensors attached to the user;

2) User-guided –when the user guides the avatar by defining its tasks and movements, and;

3) Autonomous –where the avatar has an internal state that depends on its goals and its environment, the state is directly or indirectly modified by the user and the NVC is automatically generated according to the new state [3].

As far as NVC features are automatically digitized from the user, they should be more revealing and spontaneous; but, succinct metaphors to display nonverbal cues also support the users' communication.

Now then effective collaborative learning includes both learning to collaborate and collaborating to learn, the students may require guidance in both collaboration and task oriented issues [10], while facilitating only collaboration is not particularly attached to the task at hand.

The modeled virtual agent intends is to guide collaboration for effective collaborative learning; its modeling assumes a CVE for learning in which the users' avatars interact to take care of a spatial task. The agent will not comprehend the students' dialogue; in doing so, generic analysis can be conducted and it can be mixed with other tutoring approaches like task oriented or dialogue analysis.

2 Nonverbal Communication in Collaborative Interaction

Broadly defined, nonverbal behavior might include most of what we do with our bodies; it includes also certain characteristics of verbal behavior by distinguishing the content, or meaning, of speech from paralinguistic cues such as loudness, tempo, pitch or intonation [13]. The use of certain objects like our decided outfit, or the physical environment when used to communicate something, without saying it, has also traditionally being considered as NVC.

Although NVC changes from person to person and from one culture to other, it is also functional, which means that different functional uses will lead to different arousal, cognitive and behavioral patterns of interchange [13]. Therefore, for its analysis it is particularly important taking into account its purpose.

Following Miles L. Patterson [13], the nonverbal involvement which purpose is to facilitate service or a task goal is essentially impersonal and usually constrained by the norms of the setting. NVC during collaborative interaction is more likely to have a routinely nature for interactants; gazes, pointing gestures or proximity to others will be mainly aimed to the achievement of the task.

2.1 Collaborative Learning

In the accomplishment of a task, a desired situation for an effective learning session should be a starting planning period that will help to create a shared ground or common ground [5] and to define how thinks are going to be done; followed by the implementation, that is, the task accomplishment on itself; and from time to time an evaluation episode where the students re-analyze their plans or the implementation, and change what is not appropriately working; whereas all the students have a significant participation in the three stages: i.e. planning, implementing and evaluating.

Group learning possibilities grow with its members' participation [16]. While the students participation in dialogue allows them to create a shared ground, which implies that they share knowledge, beliefs and assumptions of the task at hand in order to be able to work on it together [5]; an active student's participation corroborates that she/he is interested and understands the group activity. For collaborative learning students' participation is expected to have symmetry in both decisions making and implementing.

Jermann [10] suggested that by contrasting the students' participation in dialogue and implementation different types of division of labor can be inferred as follows:

- In sessions with *symmetric* in both *dialogue* and *implementation*: the absence of division of labor.
- The students' *symmetric* participation in *dialogue* and their *asymmetric* participation in *implementation*: a role based on division of labor without status differences, where subjects discuss plans for action together but only part of them does the implementation.
- The asymmetric participation in both dialogue and implementation: a hierarchic role

organization where some give orders and others execute them.

And the problem-solving strategies:

- Dialogue and implementation alternation could reflect a systematic problem solving approach which follows the plan-implement-evaluate phases.
- While almost null participation in dialogue and continuous implementation could reflect a brute force trial and error strategy.

Since the final purpose for the analysis of NVC cues is to model a virtual agent, the NVC cues retrieved from the environment have to be totally recognizable by a computer system.

In [14] has been argued that the collaborative interaction analysis based on NVC can be conducted with the cues available at the environment. On the other hand in a study conducted in a real life situation results showed that group member' participation rates in two NVC cues: i.e. amount of talking time and time of manipulation in the workspace, corresponded to their contribution to the accomplishment of the task; and that certain NVC cues: frequency of vocalizations, object manipulation, pointing gestures and gazes to peers can be the means to differentiate when the participants were planning, implementing or evaluating [15]. The rationalizations for the automation of these NVC cues during a collaborative learning session are next discussed:

Discussion period. During discussion episodes, planes, evaluation and agreements are settled. Then they should be distinguished from situations like a simple question-answer interchange, or the statements people working in a group produce alongside their action directed to no one in particular [8]; for that, a number of talk-turns involving most of the group members might be an appropriate method.

A talking turn, as defined by Jaffe and Feldstein [9]: begins when the student starts to speak alone and is kept while nobody else interrupts him/her. For practical effects, in a computer environment with written text communication, the talking turn can be understand as a posted message, and in oral communication to a vocalization placed by the user. Based only in talking turns, discussion periods could be inferred as in (1).

$$\text{Discussion period} \Rightarrow (\text{number of talking turns} > \text{threshold A}) \wedge (\text{number of group members involved} > \text{threshold B}) \quad \textbf{(1)}$$

Artifact manipulation. When the task at hand involves objects, their manipulation is necessarily part of the interaction; e.g. it can be the answer to an expression. The artifacts or objects related to the learning session also represent the students' shared workspace.

During the planning and reviewing phases scarce implementation should be expected, the objects will be probably more just touched than moved, while in the implementation phase there has to be significant activity in the shared workspace. The initiation of the implementation phase can be established, like with discussion periods and based only in the workspace activity, through a degree of manipulation and a number of students involved (2).

$$\text{Implementation phase} \Longrightarrow (\text{number of objects manipulated} > \text{threshold C}) \wedge (\text{number of group members involved} > \text{threshold D}) \quad \textbf{(2)}$$

When implementation is made by division of labor, this activity will probably appear in different locations at the same time. Considering, for example a group of five people, if at least two students are working in different area(s) than the other three, then division of labor could be assumed as in (3).

$$\text{Division of labor} \Rightarrow \text{number of students working in different areas of the workspace} > \text{threshold E} \quad \textbf{(3)}$$

As mentioned, a combination of amount of talk and amount of manipulation can be used to understand division of labor and the followed problem solving strategy [10].

Deictic gestures. Deictic gestures are used for pointing, they can be performed in a VE through the user's avatar body movements such as a gaze or a hand movement, but they can also be the mouse pointing.

When a group is working with objects, the communication by reference serves to get a common focus in a quick and secure form [5], [7]. Thus, the directed deictic gestures to the workspace are useful to determine whether students are talking about the task. An isolated deictic gesture could be just an instruction given or a gesture to make clear a statement about an entity, while turns of deictic gestures can be related to creating shared ground, which in turn could be related to the planning phase. In such a way that during the planning stage the students' alternation of deictic gestures and talking turn can be expected as in (4).

$$\text{Planning phase} \Longrightarrow (\mathbf{1}) \wedge (\text{alternated deictic gestures to the workspace} > \text{threshold F}) \quad \textbf{(4)}$$

Gazes. Gazes usually have a target; this target indicates the students' focus of attention. By observing the students' gazes it can be overseen if the group maintains focus on the task, and they could also be helpful to measure the students' involvement degree on it.

In order to establish if a student is involved in the task, his/her gazes should be congruent with what is going on in the environment, that is, usually gazing to the speaker or to what he/she is pointing at during a discussion period then (5), and usually to the workspace during implementation as showed in (6).

A range from 70 to 75 per cent of the time for the student to maintain this congruence is suggested as an acceptable rate in gaze behavior [1]. In that same way, it can be overseen if the group as a whole maintains focus on the task.

$$\text{Gaze target congruence in } (\mathbf{1}) \Longrightarrow \% \text{ of the gazes directed to the speaker} \vee \text{to the object pointed by the speaker} \quad \textbf{(5)}$$

$$\text{Gaze target congruence in } (\mathbf{2}) \Rightarrow \% \text{ of gazing directed to the workspace} \quad \textbf{(6)}$$

Including gazes to the analysis may provide accuracy to the distinction of a discussion period, the implementation phase and when division of labor. Where the gazing behavior expected in discussion periods could be the student's field of view directed to the peers with short shifts to the workspace and going back to peers, then (7); during the implementation stage, the student's field of view directed to the workspace with shifts to peers then (8); and in division of labor, the students working in different areas and gazing most of the time only to what they are doing as in (9).

Discussion period ⇒ (number of talking turns > threshold A) ∧ (number of group members involved > threshold B) ∧ (**5**) **(7)**

Implementation phase ⟹ (number of objects manipulated > threshold C) ∧(number of group members involved > threshold D) ∧ (**6**) **(8)**

Division of labor ⇒ number of students working in different areas of the workspace > threshold E ∧ (**6**) **(9)**

Now then the reviewing phases are expected to interrupt or to appear at the end of an implementation phase, the end or interruption of the implementation phase in the environment will be manifested as in (10). The end of the implementation phase can represent also the end of the accomplishment of the task. The reviewing phase should convey discussion periods and in some cases some workspace activity as a result of that review.

Implementation phase pause ⟹ ∃ (2) ∧ (number of objects manipulated < threshold C) ∧ (number of group members involved < threshold D) **(10)**

Reviewing phase ⟹ ∃ (**10**) ∧ (**7**) **(11)**

Adding gazes to the analysis should provide accuracy to the distinction of the reviewing phase, where the task results have to be observed in a more extended area than just an object as when implementing, the gazes will be spread in the area under review.

The statistical dispersion formula can be applied to identify the spread of gazes. Data of the gaze targets of the students collected during the implementation phase will provide their standard deviation. To quantify "nearly all" and "close to", it can be used the Chebyshev's inequality that states that no more than $1/k^2$ of the values are more than k standard deviations away from the mean to understand the spread of gazes (12). For example, for 2 standard deviations it is 1/4 = .25, then if more than 25% of the gazes are out of the range of 2 standard deviation then gazes have been spread over the workspace.

Gaze target spread in the workspace ⟹ threshold F σ of gaze targets during an implementation ≥ 1/(threshold F) 2 **(12)**

Reviewing phase ⟹ ∃ (10) ∧ (7) **(13)**

Some of these assumptions were observed in an experimental application, which results are showed in the next section.

3 Preliminary Study

A preliminary study was conducted with the purpose to understand the group process phases: planning, implementation, and evaluation, identifying patterns derived from certain NVC cues extracted from the group behavior during the session.

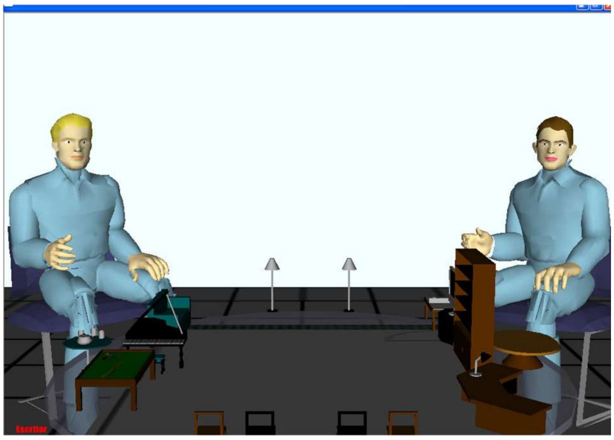

Fig. 1. Experimental application.

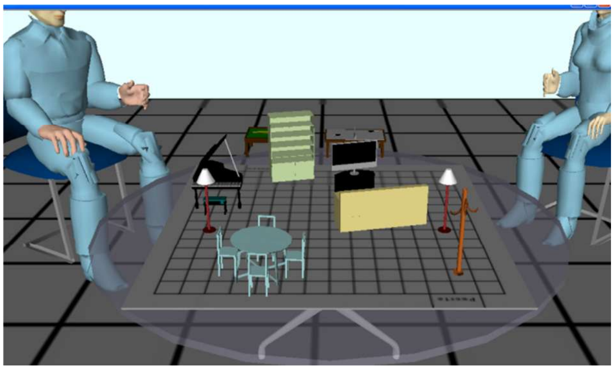

Fig. 2. Seeing down to the workspace.

The experimental application allows three users net-connected people to work in a collaborative task; the three users' avatars are placed around a table, the workspace.

The NVC cues in the environment and the students' actions available in the CVE are narrowed to those wanted to be observed and measured, avoiding other possibilities like navigation. These NVC cues are: talking turns, objects manipulation, gazes to the workspace and to peers, and pointing to objects. The avatars do not have a 'natural behavior'; they are just seated representations of the user that need a metaphorical representation of their actions in the environment. The user does not see his/her own avatar (see Figure 1).

The NVC cues are user-guided transmitted to the CVE through the keyboard and the mouse. The significant entities associated to the avatars actions are:

- Colored arrows coupled to their hair color (yellow, red, or brown) that take the place of their hands, and can be used to point the objects and/or grab them to move them;
- The avatars' head is another entity that can take four positions to change the user field of view –the change of view is controlled with the four keyboard arrows. To the front where the other two peers can be seen, to the right or left to see directly one of the peers, or down to see the workspace (see Figure 2); and,
- When the user is speaking a dialogue globe appears near his/her right hand –the user has to press the spacebar for the others to hear his/her voice.

3.1 Method

Subjects. Fifteen undergraduate students, 14 males and 1 female from the Informatics School at the Universidad of Guadalajara were invited to participate. Five groups were voluntarily formed of triads.

Materials and Task. The task consisted on the re-arrange of furniture on an apartment sketch to make room for a billiard or a ping-pong table; they decided which one of them. Sessions were audio recorder.

Procedure. A number of rules with punctuation were given regarding on how to place furniture such as the required space for the playing table, spaces between furniture and restriction on the number of times they could move furniture. Participants were allowed to try the application for a while before starting the task in order to get comfortable with its functionality. The time to accomplish the task was restricted to 15 minutes.

Data. Every student intervention within the environment was recorded in a logs file. The logs content is the user identification; the type of contribution he/she makes: i.e. move furniture, point furniture, a change in the point of view of the environment, when speaking to others; and the time the contribution was made with minutes and seconds.

3.2 Results

At a first glance to the data it could be overseen that the pointing mechanism was barely used; the speech content revealed that the users' had to make oral references to areas where there were no furniture because they could not point them.

Other identified problem related to gazes was that when the user was viewing the workspace area, he/she did not receive enough awareness about other users' gazes (see Figure 2); users had to verbally specify who they were addressing to if not to both members. We were particularly interested in observing if this unnatural gaze behavior was going to be accepted and used by the students.

Unfortunately, due to these misconceptions in the design of the environment gazes and pointing gestures had to be left out.

Discussion periods were defined as when the three group members had at least one talking turn. In order to determine the end of a discussion period, pauses of silences were considered in the range of three seconds; for automatic speech recognition, the end of an utterance is usually measured when a silence pause occurs in the range of 500 and 2000 ms [6], also the answer to a question usually goes in a smaller range, around 500 ms [19].

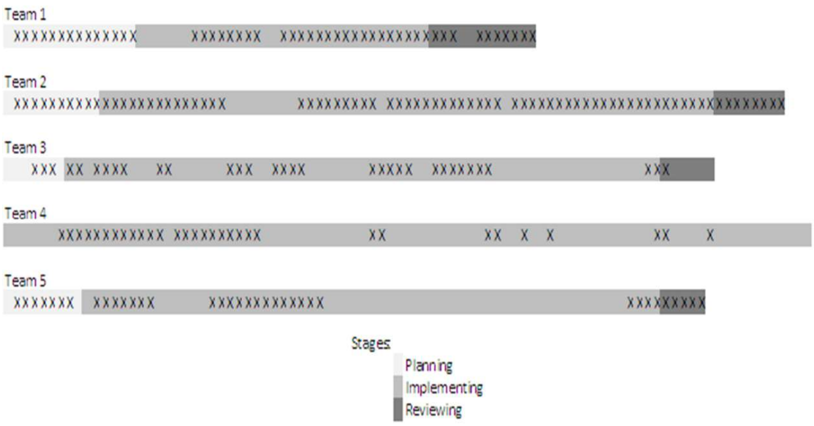

Fig. 3. The team stages during the sessions.

An external person was asked to determine through audio recorders, for each talking turn interchange whether the students were having an episode in which they were taking decisions, making plans or reviewing those; only two interchanges involving two of the three members had one of these characteristics, therefore most of the talking turn interchanges with the three members involved were discussion periods.

The stages were established as follows:

1. Planning stage – when discussion periods occur at the beginning of the session.
2. Implementation period – when at least one piece of furniture was moved.

3. Reviewing phase – discussion periods within the implementation period or at the end of it.

In Figure 3, graphics for each team session stages are presented, discussion periods are marked with an 'X'. A number of analyses can be derived from the distinction of these stages in the collaborative session such as stages times versus task effectiveness, of other related to group personality like cohesiveness.

Regarding a collaborative intelligent tutor, a clear opportunity to intervene is the fourth team which started with the implementation and then they had a discussion period while they kept the implementing, and continue working it seems that almost in silence. In the audio tape at some point they commented – "remember that we are not supposed to talk" with apparently no reason, and work to the end of the task in silence. However they fake talking, that is, they press the talking turn key probably to bring the others attention.

4 Discussion and Future Work

Nonverbal communication in a CVE could be the means to understand to a certain point what takes place during a learning session, its automatic analysis is proposed here as a tool for a virtual tutor to guide students to enhanced collaboration, as when the students are expected to start with a planning stage in which they create a common ground, and with an implementation that includes reviewing periods.

In order to understand these assumptions an experimental application was used to conduct a preliminary study. Unfortunately two misconceptions on its design invalidate the gazes and pointing mechanisms. Although, results showed that planning, implementing and reviewing stages can be distinguished through the retrieval from the logs of the talking turns and the manipulation of objects, also discussion periods can be determine.

There is no doubt of the importance of awareness; the lack of awareness for the other users invalidates the visual advantages in CVEs. An obvious next step is to adapt the application to give feedback for gazes when users have the view towards the workspace; along with a more free pointing facilitation to the entire workspace.

In this paper assessment rules for the automatic analysis of NVC cues were presented, using these methods a virtual tutor to facilitate collaborative interaction in a learning scenario can be modeled.

References

1. Argyle, M., & Cook, M.: Gaze and mutual gaze. Cambridge University Press, Cambridge (1976)

2. Bricken, M.: Virtual Reality Learning Environments: Potentials and Challenges. Computer Graphics, 25, 178–184 (1991)
3. Capin, T. K., Pandzic, I. S., Thalmann, N. M. et al.: Realistic Avatars and Autonomous Virtual Humans in VLNET Networked Virtual Environments (1997)
4. Chittaro, L., & Ranon, R.: Web3D Technologies in Learning, Education and Training: Motivations, Issues, Opportunities. Computers & Education Journal, 49, 3-18 (2007)
5. Clark, H. H., & Brennan, S. E.: Grounding in communication. In: Resnick, L.B., Levine, J.M. and Teasley, S.D. (eds.) Perspectives on socially shared cognition, pp. 127–149. American Psychological Association, Hyattsville, MD (1991)
6. Edlund, J., Heldner, M., Gustafson, J.: Utterance Segmentation and Turn-Taking in Spoken Dialogue Systems. Computer Studies in Language and Speech, 8, 576–587 (2005)
7. Gergle, D., Kraut, R. E., Fussell, S. R.: Language Efficiency and Visual Technology: Minimizing Collaborative Effort with Visual Information. Journal of Language and Social Psychology, 491–517 (2004)
8. Heath, C., Jirotka, M., Luff, P. et al.: Unpacking Collaboration: The Interactional Organisation of Trading in a City Dealing Room. Computer Supported Cooperative Work, 3, 147–165 (1995)
9. Jaffe, J., & Feldstein, S.: Rhythms of dialogue. Academic, New York, NY (1970)
10. Jermann, P.: Computer Support for Interaction Regulation in Collaborative Problem-Solving (2004)
11. Jermann, P., Soller, A., Lesgold, A.: Computer Software Support for Collaborative Learning. In: Anonymous What We Know About CSCL in Higher Education, Kluwer, Amsterdam pp. 141–166 (2004)
12. Martínez, A., Dimitriadis, Y., de la Fuente, P.: An XML-Based Model for the Representation of Collaborative Action (2002)
13. Patterson, M. L.: Nonverbal Behavior: A Functional Perspective. Springer-Verlag, New York (1983)
14. Peña, A., & de Antonio, A.: Inferring Interaction to Support Collaborative Learning in 3D Virtual Environments through the User's Avatar Nonverbal Communication. IJTEL, 2, 75–90 (2010)
15. Peña, A., & de Antonio, A.: Nonverbal Communication as a means to support collaborative interaction assessment in 3D Virtual Environments for learning. In: Juan, A.A., Daradoumis, T., Xhafa, F., et al (eds.) Monitoring and Assessment in Online Collaborative Environments: Emergent Computational Technologies for E-learning Support,IGI-Global, Hershey, PA, pp. 172–197 (2009)
16. Scrimshaw, P.: Cooperative writing with computers. In: Scrimshaw, P. (ed.) Language, classrooms & computers, pp. 100-110. Routledge, London-UK (1993)
17. Soller, A., Jermann, P., Muehlenbrock, M. et al.: Designing Computational Models of Collaborative Learning Interaction: Introduction to the Workshop Proceeding (2004)
18. Sonnet, H., Carpendale, S., Strothotte, T.: Integrating Expanding Annotations with a 3D Explosion Probe, 63–70 (2004)
19. Stivers, T., Enfield, N. J., Brown, P. et al.: Universals and Cultural Variation in Turn-Taking in Conversation. PNAS, 106, 10587–10592 (2009)
20. Winn, W. D.: Current Trends in Educational Technology Research: The Study of Learning Environments. Educational Psychology Review, 14, 331–351 (2002)
21. Wolff, R., Roberts, D., Steed, A. et al.: A Review of Tele-Collaboration Technologies with Respect to Closely Coupled Collaboration. IJCAT (2005)