Jl. of Interactive Learning Research (2007) **18***(2)*, 231-250

Beyond Logging of Fingertip Actions: Analysis of Collaborative Learning Using Multiple Sources of Data

NIKOLAOS AVOURIS, GEORGIOS FIOTAKIS,
GEORGIOS KAHRIMANIS, MELETIS MARGARITIS,
AND VASSILIS KOMIS
University of Patras, Greece
avouris@upatras.gr

In this article, we discuss key requirements for collecting behavioural data concerning technology-supported collaborative learning activities. It is argued that the common practice of analysis of computer generated log files of user interactions with software tools is not enough for building a thorough view of the activity. Instead, more contextual information is needed to be captured in multiple media like video, audio files, and snapshots, in order to re-construct the learning process. A software environment, Collaborative Analysis Tool, (ColAT) that supports interrelation of such resources in order to analyse the collected evidence and produce interpretative views of the activity is described.

Introduction

Collection of usage data by registering users' operations in the form of log files has become mundane during technology-supported learning activities these days. Many researchers assume that learning and cognitive processes can, in principle, be inferred from studying and analysing this recorded behaviour (Hulshof, 2004). Logfile analysis can be used when the purpose is to infer the cognitive processes and social behaviour of persons who interact with software tools. Subsequently, analysis can be performed in a number of ways, for example by examining the frequency with which different operations are carried out or by focusing on the sequence in which operations occur. Analysis of a learning activity is important for understanding the complex process involved, improving effectiveness of collaborative learning approaches and can be used as a reflection-support mechanism for the actors involved.

Tools to support interaction and collaboration analysis have been proposed in the field of learning technology design and human-computer interaction (Dix, Finlay, Abowd, & Beale, 2004). In the education field, analysis of collaboration and interaction between the actors, such as students, tutors, the artefacts and the environment is a process that can support understanding of learning, evaluate the educational result and support design of effective technology (Gassner, Jansen, Harrer, Herrmann & Hoppe, 2003). Many researchers have studied the problem of combining multiple sources of data during interaction analysis. For example, Heraud, Marty, France and Carron (2005) proposed combination of keystroke log files and web logs. However, the more challenging question, discussed in this article, is to combine structured data, like log files with unstructured ones, like audio and video recordings in the same environment.

In this article, we describe first the typical characteristics of a software environment that records users' operations and then supports their analysis during the activity and off line. In the second part of the article, we argue further that while this approach is useful, more contextual information is needed to be interrelated to the collected log files. So an innovative analysis tool (ColAT) is presented that can be used for effective analysis of interrelated multiple data that may be collected during technology-supported learning activities.

Logfile-based Analysis of Learning Activities

One of the new opportunities that information and communication technologies offer to learning activities is related to automatic logging of actions by the computer environments used. The outcome of this process, in the form of a log file, may be used for analysing and evaluating learning activities. Evaluation can then lead to improvement of learning practices and the computer tools used.

A suitable field for the application of log file analysis is Computer-Supported Collaborative Learning (CSCL). Evaluation of individual computer-supported learning activities often involves comparisons of pre and post tests indicating levels of knowledge of students. What is assumed by this practice is that learning activities cause individual cognitive processes that are not accessible per se, but only through their outcomes. On the contrary, during collaborative learning social interaction is added to learning activity, so what one participant communicates with others is accessible to researchers, facilitating analysis of the learning process (Dillenbourg, Baker, Blaye & Malley, 1996). The computer is often used as a tool facilitating peer interaction and communication, thus a record of social activity is added to that of interaction with learning content or problem solving operations. The state of evolving knowledge must be continuously displayed in this case by the collaboration participants with each other (Stahl, 2001). So logging and analysing of user-computer tool interactions is of added value when referring to CSCL.

There are many different approaches to log file analysis, especially in the case of collaborative activities. In the next section, some of them are presented through a collaborative problem solving environment that integrates a wide range of log file analysis tools.

Logfile-based Analysis with the Use of a CSCL Environment

In this section, we describe the functionality of a typical environment for analysis of group learning, called Synergo (www.synergo.gr), associated to a synchronous collaboration-support environment, which permits direct communication and problem solving activity of a group of distant students, manipulating a shared graphical representation (Avouris, Margaritis & Komis, 2004). Synergo keeps track of user operations. It also incorporates tools for analysis of these usage log files. Through them, the researcher can play back the recorded activity offline and annotate the jointly produced problem solution, usually in a graphical form (e.g., a concept map, a flow chart etc.), while various indicators and views of the log files can be produced.

In a typical synchronous collaborative learning situation in which Synergo is used, two or more actors, supported by networked equipment, collaborate at a distance by communicating directly though an integrated chat tool and by acting in a shared activity space. A graphic representation of a solution to a given problem appears in this shared activity space. This activity is typically tracked through logging of the main events of the actors in the shared space and of the text dialogue events.

The Synergo analysis tools are used for presentation and processing mainly of these log files, produced during collaborative learning activities. These log files (see an example at the top in Figure 1) contain time-stamped events, which concern actions and exchanged text messages of partners engaged in the activity, in sequential order.

These events have the following structure:

{<ID>, <time-stamps>, <actor>, <event-type>, <attributes>, <comments>}.

Some of these fields take their value automatically by the Synergo software. An example from the log file of Figure 1 is the following: {ID = 623, Time1 = 00:18:11, Time2 = 02:02:28, User = hlias, Action = "Insert Concept Relationship," Attributes = "qualitative(57), x=320, y=304" }. This is a record of an event produced at 00:18:11, that occurred 02:02:28 since the beginning of the activity (relative time), by user *Hlias* who inserted in the shared activity space an object at position x=320, y=304.

Some more attributes can be associated to the log file records. The <event type> attribute categorizes the recorded event. This categorization can be

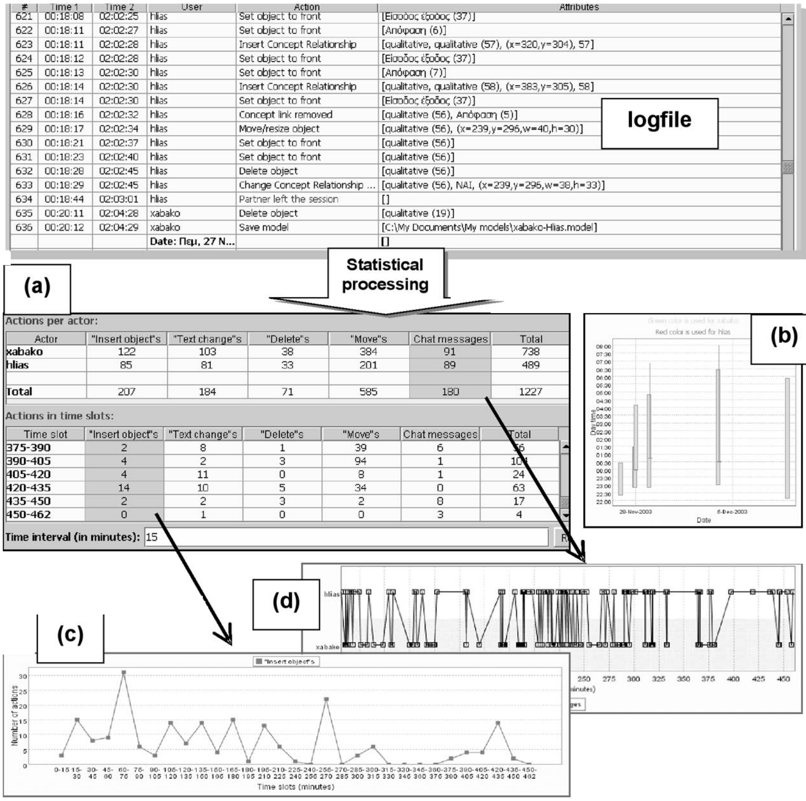

Figure 1. Synergo analysis tools: The log file (top of the picture) is processed for producing statistical indicators across various dimensions (type of event, time slot, actor), shown in (a). Also, the group sessions over time are shown in (b), while in (c) and (d) the statistical indicators are plotted vs. time

done by interpreting one by one the log file events manually. The Synergo environment facilitates this tedious process, by allowing association of kinds of events, automatically generated by the software, to classes. So for instance, all events of type "*Change of textual description of concepts*" in a concept-mapping tool are associated to the "*Modification*" general type of action, as shown in Figure 2.

Following this first level of automatic annotation of the log file, statistics and visual views concerning the activity can be automatically generated. For instance, in Figure 1 some of the views automatically generated by the Synergo analysis tools can be seen. This is an extract from a log file that was

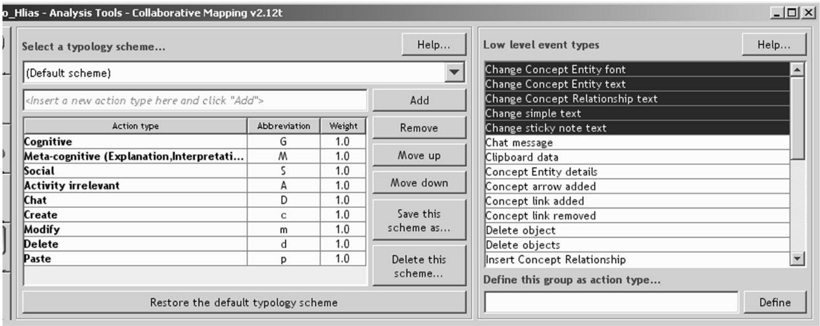

Figure 2. Definition of an Event Typology scheme: The low-level recorded events, generated by the software (left) are grouped to action types (right)

generated by a pair of two students of a distance learning course who interacted for over seven hours (462 minutes of interaction spread in 8 sessions). In Figure 1(a) the recorded events are grouped by user and type of event in the top table and by time interval and type of event in the second. The analyst can observe the value of various indicators, like the number of events of type "insert new object in the activity space" per time interval, shown in Figure 1(c), or an interaction diagram indicating the activity per partner of a specific type of event, like chat messages between two partners in Figure 1(d). Finally, a view related to length of sessions in Figure 1(b). These representations can have some value for a trained analyst or teacher, or they can be used as self-awareness mechanisms for students as they can be presented to them during collaborative activities.

Not all recorded events however can be automatically annotated in this way, while important events are not captured at all by the log file, as they do not occur as a result of user-tool interaction (i.e., user fingertips activity). For instance, face to face dialogues have to be captured through other media, and interpreted by the analyst. So, after establishing their meaning and intention of the interlocutor, may be annotated accordingly. There are various ways of interaction, for instance, a suggestion of a student on modification of part of the solution can be done either through verbal interaction or through direct manipulation of the objects concerned in the shared activity space.

In addition, more complex indicators may be generated. An example is the graph of evolution of the *Collaboration Factor*, discussed in (Avouris, Margaritis, & Komis, 2004). This index reflects the degree of contribution of actors in the solution of a problem solving task, taking into account the relative weights of actors, components of the solution and types of actions. The *Collaboration Activity Function* (Fesakis, Petrou & Dimitracopoulou, 2003), constitutes a similar index that calculates the value for collaboration by taking into consideration the actions performed by users in collaborative environments through all collaboration channels (e.g., shared workspace and chat). In larger group settings, *Sociograms*, meaning graphic representations of the social links between students, based on the quality and quantity of interactions between them, may be used for representing group relations (Reffay & Chanier, 2003).

In general, it has been observed that many log files, like the Synergo log file presented in this section, bear many similarities, both in synchronous and asynchronous collaboration support environments. So it is possible to define a common format and an ontology for representing the semantics of collaborative learning log file data (e.g., Kahrimanis, Papasalouros, Avouris & Retalis, 2006), thus facilitating exchange and interpretation of log files by various researchers. However, despite the increased popularity of log file-based analysis of learning activities, and the useful data and views generated from them, these views may not be enough for gaining a full understanding of the activity and may lead to false interpretations and conclusions. In the following section, the main concerns and shortcomings of analyses restrained to just log file data are reported. First the limitations that are due to loss of information conveyed through additional communication channels in collocated and distant settings are discussed, followed by the specific requirements of mobile learning situations.

Shortcomings of the Log File Analysis Approach

Computer-supported collaborative activities in the simplest level are classified according to two dimensions: a spatial and a temporal one. In the spatial axis, collaboration activities are discriminated between collocated and distant ones. In the temporal axis. the distinction refers to synchronous and asynchronous activities. Logfile analysis is not favoured equally in all modes of communication, as discussed in the following.

The Case of Collocated CSCL Activities

The most problematic cases of use of log files as the only input to analysis are collocated collaboration activities. In such activities, a computer tool used constitutes just one of many different communication channels. The fact is that such a setting does not inhibit oral communication, even if the tool used provides support for exchanging text messages. Furthermore, secondary channels of face to face communication may convey important meaning to the analysis of the activity. Gestures, intonation, facial expressions and posture of students should not be neglected in such case. Moreover, the structure of oral dialogues, in contrast to typed messages, is not easily defined. Important information that has to be considered refers to turn-taking, overlapping, hesitation of one partner or intervals of silence.

When students are expected to collaborate face to face, the inadequacy of log files for analysis is rather obvious. However, there are cases, (e.g. in a computer equipped classroom), where students are not supposed to have straight face to face communication, that they actually do so. For example, a CSCL environment like Synergo may be used, which provides a synchronous chatting tool. According to our experience, it is not unlikely that collaborating students occasionally engage themselves in oral dialogues during problem solving activity, even if they have to move from their workstations. Such cases may be tricky for an analyst, because the bulk of communication is conveyed through the CSCL tool, but important information communicated orally may escape their attention.

The Case of Distant CSCL Activities – An Example

In distant CSCL activities, researchers and activity designers often seem to have a misleading perception of the nature of CSCL activities. They sometimes develop strict educational scripts, provide certain CSCL tools to the students and restrict the students to conduct an activity according to the given directives. However, in practice, students prove to be surprisingly flexible in terms of usage of computer tools – they adopt alternative media in order to interact with their peers. Usage of email, instant messengers and asynchronous discussion forums are the most common examples. The fact that researchers, in contrast to face to face collaboration, cannot physically observe interactions may lead them to completely ignore such practices.

An experience related to such practices is reported in a cross-national CSCL activity between Greek and German universities that provides an example of both synchronous and asynchronous collaboration (Harrer, Kahrimanis, Zeini, Bollen & Avouris, 2006). Students from both universities were assigned a task as homework. They were requested to work in dyads with a distant partner, using provided collaboration support tools. In addition, an asynchronous discussion forum was set up, so that students could exchange messages for knowing each other better and planning their work. Students were asked to deliver a report on the activity containing data from any tools used, in order to demonstrate their collaborative activity.

This scenario left a lot of freedom to the students to approach their task, in terms of when and how to work together or how to divide the work. The facilitators of the activity, who were researchers aiming to study this kind of distant collaboration activity, preferred to give such freedom to students instead of setting a more contained lab situation. The latter case might have been preferable for controlled analysis of some collaboration aspects, but would have produced an artificial environment that would not have been connected well to the students' real-world experiences.

The reports gathered at the end of the activity revealed that most pairs used additional tools in order to communicate. Five out of ten groups used an instant messenger application and 50% of the groups exchanged email messages. This was rather surprising having in mind that the students had plenty of tools at their disposal. The recommended collaboration support environment contained integrated chat tools for synchronous communication and a forum for asynchronous communication was also provided.

Many students negotiated parts of the problem through chat messages conveyed through external chatting tools and then used the collaboration support environments to finalise the problem solution. Others worked on their own and sent a proposal of a solution to their partners by email.

Ignoring that students used other tools than the suggested ones, or underestimating the importance of information conveyed through them would restrain a researcher from understanding thoroughly the studied activities. However, even if one is aware of that problem, it is impossible to gather all data of student communication. In addition to the technical problems, it is expected that students would not always be willing to report them to their supervisors for privacy reasons.

Moreover, even if one manages to gather all logged data (regardless of the tool that produces them), that may still not be enough to gain a thorough view on the activity. Students may consult external resources while collaborating (e.g., books, the web) in order to find information. They may also get themselves involved in individual tasks that help them learn. No information on such individual activities can be gained by any kind of log files. In the study reported here, it was found that in many cases students worked on their own for some time and then they were involved in collaborative sessions. In the beginning of these sessions, they negotiated their individually produced partial solutions of the problem. That is a general problem when analysing collaborative activities and especially asynchronous ones. Not all knowledge gained is a product of collaboration. In most cases, collaborative sessions interplay with individual learning, leading to learning results that cannot be easily attributed to one practice or the other.

Requirements of Mobile Learning Activities – An Example

In the last years, collaborative learning practice favours the use of handheld devices. Future classrooms are likely to be organized around wireless Internet learning devices that will enable a transition from occasional, supplemental use of learning technology in real-world education to frequent, integral use (Roschelle & Pea, 2002). This constitutes a major shift in CSCL practice according to many perspectives. First, a wide range of different sources of information and knowledge may be available for students participating in the same activity. Control over the software used and the modes of communication between students would be very difficult. Moreover, the way that multiple sources of knowledge interplay would not be easily determinable. Adding to the above, the use of peer-to-peer communication architectures that are more likely to prevail in handheld device interactions, the desire of logging all data and integrating them would be rather unrealistic.

In addition, when analyzing such cases one has to face the same problems as with classic face-to-face collaborative activities. The above reasons justify the claim that analysis of log files of use of handheld devices is inadequate for a thorough analysis of mobile learning activities.

In order to give a simple example for such limitations, we describe the experience of designing a collaborative learning activity for a traditional historical/cultural museum (Cabrera et al., 2005). The activity, based on a "Mystery in the Museum" story, involves collaboration of small groups of students through mobile handheld devices. An application has been built that permits authoring of such activities, while a usability evaluation study was performed that revealed some of the limitations of the design.

The plot involved a number of puzzles that relate to the exhibits of the museum and their solution brings rewards to the players. These puzzles, the most typical examples of which involved scrabbled images of certain exhibits and verses found in manuscripts of the museum, necessitate collaboration for their solution, as the necessary pieces were spread in the mobile devices of the members of the group (see Figure 3). A negotiation phase was initiated then that resulted in exchange of items that could lead the group to a solution of the particular puzzle. The rewards had the form of clues that help the players solve the mystery. Since a large number of children (e.g., a school party) may be organized in multiple groups, the intention was to create competition among different groups. The aim of the activity was to mix the real and the virtual world and to make children work together in a collaborative way in this setting.

To move from evaluation of the technology used to evaluation of collaborative learning, log file analysis cannot offer much. Table 1 summarizes calculations based on action logs, as reported by (Stoica, Fiotakis, Simarro, Muñoz Frutos & Avouris, 2005). Such measures offer just indications of extreme cases of failure, like the unwillingness to work on the task.

However, no significant findings can be deduced by such measures. In a later section, we present an alternative approach to analysis that helps shedding light into cases like this.

Methodological Concerns

A serious shortcoming of log file analysis concerns the interpretation of the meaning of the unit of analysis and of the values of quantitative indicators. For instance, some chat messages logged by a tool, used in a CSCL activity, may be unimportant although they are annotated according to a coding schema and counted in certain indicators. Moreover, action counts may include routine actions as well as crucial ones that are weighted equally. Such issues reveal that quantitative measures using log file events have little reliability if they aim to test hypotheses based on assumptions of meaning of certain logged actions. Therefore, the recommended methodologies for CSCL activities analysis are mostly of qualitative nature, based on unstructured data, discussed in the following.

Figure 3. The screenshots of the handhelds of two partners during the puzzle activity

Table 1

Statistics of Logged Actions for Three Groups, G= Group ID, P= Profile (task)

G*	P**	Member	ImageGame				
			Time	Events	Exchanges		Completed
1	1	PDA 1	6'35"	101	9	3	Yes
		PDA 2	6'18"			6	Yes
	2	PDA 3	3'05"	47	3	0	Yes
		PDA 4	3'31"			3	Yes
2	1	PDA 1	4'22"	46 ~	3	2	Yes
		PDA 2	6'26"			1	Yes
	2	PDA 3	2'49"	41	5	4	Yes
		PDA 4	2'56"			1	Yes
3	1	PDA 1	2'29"	47	4	3	Yes
		PDA 2	3'32"			1	Yes
	2	PDA 3	4'59"	71	7	4	Yes
		PDA 4	4'14"			3	Yes
		Average:	4'16"	~59			
Average by profile:		1	4'57"	~65	~5		
		2	3'36"	53	5		
* Group, ** Profile							

Analysis of Computer-Supported Collaborative Learning activities constitutes a research field that bears many methodological similarities with other domains of computer-aided learning. As stated above, what is learned by one participant has to be communicated to others, providing valuable information to researchers. The core object of research is interpretation of collaborative interactions.

For this purpose, methods from the fields of ethnomethodology (Garfinkel, 1967), conversation analysis (Edwards, & Potter, 1992), interaction analysis (Jordan, & Henderson, 1995), video analysis (Heath, 1986) and ethnography (Hammersley, 1992) are applied. Most of these methodologies demand that the researchers are immersed in the culture of the students and stress the determinant role that the context plays in the learning activity.

For analysis of the activities, in addition to log files, other sources of data should be available to researchers. Video captures is one of the most important ones. Furthermore, observation notes, audio recordings and snapshots may be useful. In order not to lose the benefits that log file data provide for analysis, but to overcome the limitation of this approach as well, in the next section we propose an alternative method of analysis with the aid of an innovative analysis tool.

Interrelation of the Log File to Other Behavioural Data in ColAT

It should be observed that structured data, like a typical log file, takes usually the form of an ordered list of events occurred at the user interface of a software tool. It contains a record of the activity of one or more learning actors, from the rather restrictive point of view of their fingertip actions. However a lot of contextual information relating to the activity, as well as results of the activity in print or other forms, oral communication among the actors, is not captured through this medium. So in this section we present an analysis environment that permits integration of multiple media collected during learning activities and allows the application of qualitative methodologies discussed in the *Methodological Concerns* section of this article.

The *Collaboration Analysis Tool (ColAT)* is the environment that is used for building an interpretative model of the activity in the form of a multilevel structure, following an Activity Theory approach (Bertelsen & Bodker, 2003), incorporating pointers and viewers of various media. ColAT permits fusion of multiple data by interrelating them through the concept of the universal activity time. Figure 4 shows an example of creation of a new analysis project and interrelation of multiple sources of data. The analysis process during this phase, involves interpretation and annotation of the collected data, which takes the form of a multilevel description of the activity.

The ColAT tool, discussed in more detail in (Avouris, Komis, Margaritis & Fiotakis, 2004), uses the form of a theatre's scene, in which one can observe the activity by following the plot from various standpoints. The

Operations view permits study of the details of action and interaction, as recorded by a log file, while other media like most typically video and audio recordings, capture dialogues, other behavioural data of actors (posture, gestures, facial expressions, etc.), while media like screen snapshots and PDF files record intermediate or final outcomes of the activity. The automatically generated log of behavioural data can be expanded in two ways:

- First by introducing additional events as they are identified in the video and other media, and by associating comments and static files (results, screen snapshots, etc.) to specific time stamped events.
- Second, more abstract interpretative views of the activity may be produced: the *Actions view* permits study of purposeful sequences of actions, while the *Activity view* interprets the activity at the strategic and motivational level, where most probably decisions on collaboration and interleaving of various activities are more clearly depicted.

This three-level model is built gradually. The first level, the *Operations level*, is directly associated to log files of the main events, produced and annotated, and is related through the time stamps to the media like video. The second level describes *Actions* at the actor or group level, while the third level is concerned with *motives* of either individual actors or the group.

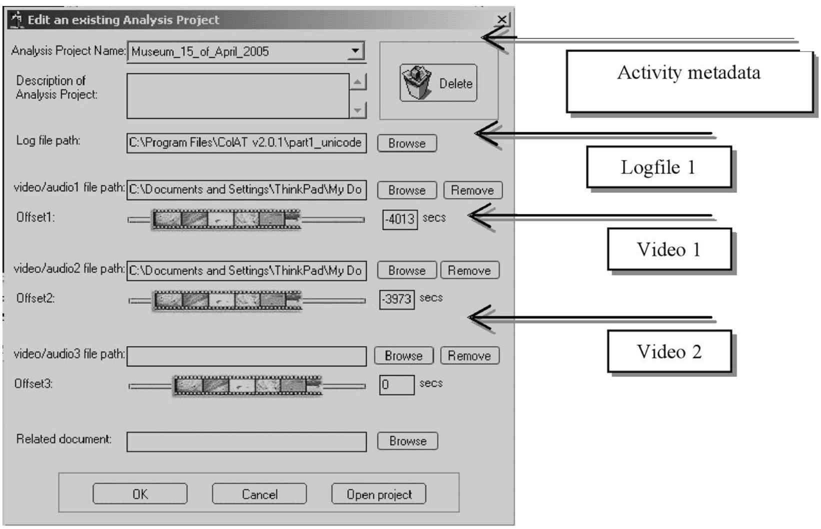

Figure 4. The ColAT environment: Project definition in which multiple log files and video/audio sources are synchronized by defining their corresponding time offsets

In Figure 5, the typical environment of the ColAT tool for creation and navigation of a multi-level annotation and the associated media is shown. The three-level model, discussed in more detail in the following, is shown on the right side of the screen, while the video/audio window is shown on the left-hand side. One other feature shown in Figure 5 is the viewer filter, through which a subset of the activity can be presented, related to specific *actors, tools* or *types of events*. So for example, the log file events related to a specific actor may be shown, or actions related to a specific tool, or a specific kind of operations.

A more detailed description of the multilevel representation of the activity shown in Figure 5 is provided next. The original sequence of events contained in the log file is shown as level 1 (*Operations level*) of this multilevel model. The format of events of this level, in XML, is that produced by Synergo, ModellingSpace, CollaborativeMuseumActivity and other tools that adhere to this data interchange format (Kahrimanis et al. 2006). Thus the output of these environments can feed into ColAT, as first level structure. A number of such events can be associated to an entry at the Actions level 2. Such an entry can have the following structure: {<ID>, <time-span>, <entry_type>, <actor>, <comment >} where <ID> is a unique identity of the Action, <time-span> is the period of time during which the action took place, <type> is a classification of the entry according to a typology, defined by the researcher, followed by the <actor> that participated in the activity, a

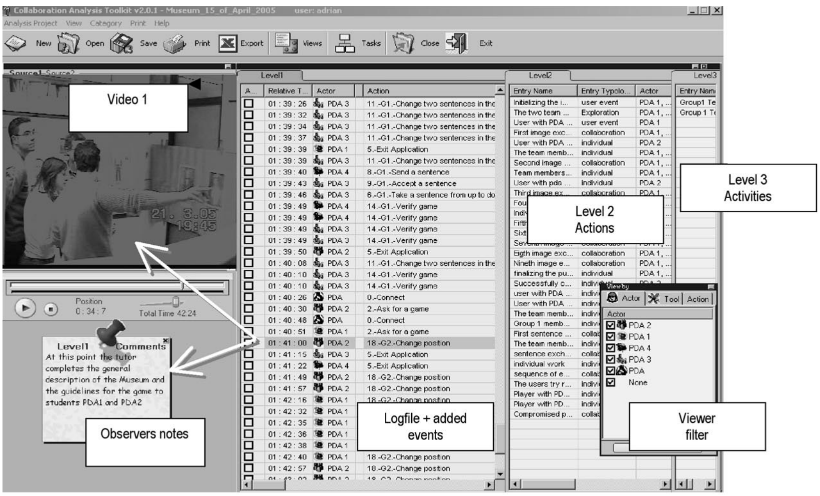

Figure 5. The ColAT environment: Multi-level view of problem solving activity, (The extract is from the study of Learning Activities in a Museum, discussed in this article and in Cabrera et al, 2005)

textual comment or attributes that are relevant to this type of action entry. Examples of entries of this level are: "Actor X inserts a link," or "Actor Y contests the statement of Actor Z."

In a similar manner, the entries of the third level (*Activity level*) are also created. These are associated to entries of the previous *Actions level 2*. The entries of this level describe the activity at the strategy level as a sequence of interrelated goals of the actors involved or jointly decided. This is an appropriate level for description of plans, from which coordinated and collaborative activity patterns may emerge. In each of these three levels, a different event typology for annotation of the entries may be defined. This may relate to the domain of observed activity or the analysis framework used. For entries of level 1 the Object-oriented Collaboration Analysis Framework (OCAF) event typology (Avouris, Dimitracopoulou & Komis, 2003) has been used, while for the action and activity level different annotations have been proposed. In Figure 6, the tools for definition of annotation scheme for actions and identity of actors and tools in ColAT is shown.

The various no-structured media, like video or audio that can be associated to logged events through ColAT can be played from any level of this multi-level model of the activity. As a result, the analyst can decide to view the activity from any level of abstraction he/she wishes, for example to play back the activity by driving a video stream from the *operations, actions* or the *activity* level. This way the developed model of the activity is directly related to the observed field events, or their interpretation.

Other media, like still snapshots of the activity or of a solution built for a given problem, may also be associated to this multilevel model. Any such image may be associated through a timestamp to a point in time, or a time interval, for which this image is valid. Any time the analyst requests playback of relevant sequence of events, the still images appear in the relative window. This facility may be used to show the environment of various distributed users during collaboration, as well as tools and other artefacts used. Also observer comments related to events can be inserted and shown in the

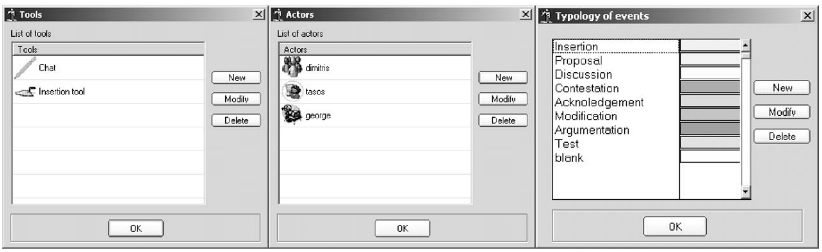

Figure 6. Definition of (a) tools used, (b) actors, and (c) typology of events relating each type of event to a specific color code, in ColAT

relevant window, as shown in the bottom left corner of Figure 5.

The possibility of viewing a process using various media (video, audio, text, log files, still images), from various levels of abstraction (operation, action, activity), is an innovative approach. It combines in a single environment the hierarchical analysis of a collaborative activity, as proposed by Activity Theory, to the sequential character of behavioural data.

Validation Studies

The discussed tools have been used in a number of studies that involved effective analysis of collected evidence of technology-supported learning activities in various forms. Three such studies are briefly presented here.

In the study reported in (Fidas, Komis, Tzanavaris & Avouris, 2005), data were collected of groups of students (15-16 years old) of a Technical Lyceum, interacting through ModelsCreator3, a collaborative modelling environment. Interaction between distant group members was mediated by a chat tool while interaction between group members that were located in front of the same workstation was mainly direct conversation. Interaction in the first case was captured through the ModelsCreator3 log file that conforms to the ColAT format, while the latter was captured through audio recording. By associating the two data sources, valuable information on comparison of the content of interaction that was done through the network and the dialogues of the group members was performed. The educational process was thus discussed according to various dimensions, like group synthesis, task control, content of communication, roles of the students and the effect of the tools used. In these studies, various features of the analysis tools presented here have been used. First, tools have been used for playback and annotation of the activity. Subsequently, the audio and sequences of still images, along with the log files of the studies were inserted in the ColAT environment through which the goal structures of the activities were constructed and studied.

In (Voyiatzaki, Christakoudis, Margaritis & Avouris, 2004) a study is discussed of activities that took place in a computer lab of a Junior High school, using the collaboration environment, Synergo. The activity involved exploration by pairs of pupils of a simple algorithm flow chart and negotiation of its correctness through the chat tool. The log files of Synergo were analysed along with contextual information in the form of video recording of the whole classroom during the activity and with observers' notes. These two data sources where interrelated and through this process the verbal interventions of the tutor where identified and the effect of these on the students problem solving process was studied. This study identified the patterns of pupils' reactions to tutoring activity.

Finally, in a third case, the collaborative learning activity about a mystery play in a museum using PDAs has been studied (Cabrera et al., 2005; Stoica, et al, 2005). In the study, a log file of the museum server was studied in relation to three streams of video from different angles together with the observers' notes. It was found that various events related to interaction of the students with the exhibits and verbal interactions of the students between them and with their tutor/guide were captured in the video streams and were interrelated with actions at the user interface level of the various PDAs that were automatically recorded by the software application used. In this particular study it was found that the additional information conveyed through the posture of the users and their spatial location was important for studying and understanding the activity, while the limited size of the portable devices and the technical limitations of monitoring the PDA screens during the activity made the video streams and interrelated logged events at the side of the server most valuable source of information.

A summary of the presented and briefly discussed studies is included in Table 2. In the three studies, the common characteristic was that in order to

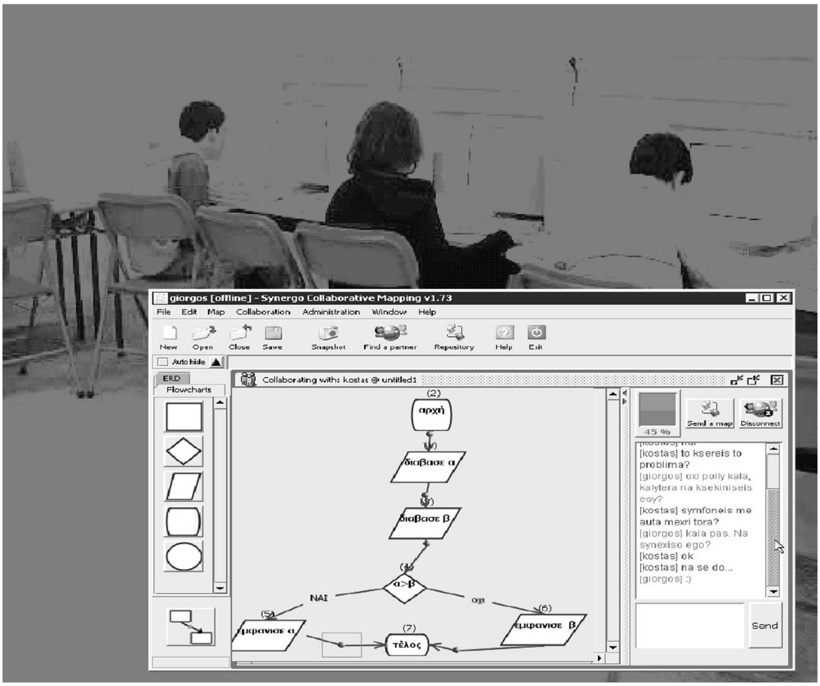

Figure 7. A view of the lab and a snapshot of a pupil workstation during the activity of the study reported by Voyiatzaki et al. (2005). The pupils in pairs had to explore a simple algorithm flow chart and negotiate its correctness, through the chat tool

Table 2
Summary of the Presented Case Studies

Study	Setting	Data Sources	Mode of collaboration	Use of ColAT
Fidas et al. 2005	Technical Lyceum, Information Technology class (15-16 year old), 20 pupils	Logfiles Observer notes audio	ModelsCreator3 throughthe network, and face to face	Interrelation of computer based activity and recorded face to face interaction, patterns of collaboration emerged
Voyiatzaki et al. 2004	Junior High School, Computer Lab (14-15 year old), 20 Pupils	Logfiles Video Observer notes Activity sheets	Synergo through the network, with tutor intervention	The teacher intervention was recorded in video and the effect on students activity was identified
Stoica et al. 2005, Cabrera et al. 2005	Historical/Cultural Museum activity School party (15 year old), 12 pupils	Logfiles 3 Video streams Observer notes	Face to face, Using wireless network-enabled PDAs	Students gestures, posture and face to face interaction captured on video and interrelated to logs of PDAs and screenshots

analyse effectively the studied activities and test their hypotheses the analysts used additional evidence in various forms, mostly video and audio. These were added to log files generated by the software tools used (chat messages exchanged, actions on concept mapping tools, etc.) and were interrelated to them. The analysis environment ColAT that was used in these cases facilitated and effectively supported the analysis and evaluation task, as described in more detail in the three study reports (Fidas et al., 2005; Voyiatzaki et al. 2004; Stoica et al., 2005)

CONCLUSIONS

In this article, we discussed the limitations of the current practice of analysis of log files and the need for using multiple sources of data during the study of collaborative learning activities. Firstly, a typical case of log file-based analysis was presented using the Synergo Analysis tool as an example. Subsequently, the limitations of such an approach were discussed in particular with relation to the requirements of collocated activities, distant collaboration activities and activities related to the use of handheld devices. Finally, the Collaboration Analysis Tool (ColAT) that permits fusion and interrelation of multiple sources of data of collaborative activities was presented and examples of its validation studies were discussed.

The log file analysis approach used as main source of data the log files of events generated by user operations in a Collaborative Learning environment, like Synergo. In this case playback and statistical indicators visualisation were used in order to re-construct the problem solution and view the partners' contribution in the activity space. However it was found that often such an approach is not adequate for a complete reconstruction of the learning activity, as essential contextual information, beyond the user fingertips actions was missing.

The second approach, involves multiple interrelated sources of data. It also involves building of a multilevel interpretation of the solution, starting from the observable events, leading to the cognitive level. This is done by using a combination of multiple media views of the activity. Through this, a more abstract description of the activity can be produced and analysed at the individual as well as the group level.

It should be observed that the two presented approaches are complementary in nature; the first one, used for building a quantitative view of the problem solving at the user interface level, while the second one leading to more interpretative structures, as it takes into account additional contextual information in the form of various other media. The result of the first phase can feed the second one, in which case the annotated log file is just one source of information. The two presented tools are quite independent, since their use depends on the available data. The Synergo Analysis Tool is mostly related to the Synergo synchronous problem-solving environment, while the ColAT tool is more generic and can be used for studying any kind of learning activity, which has been recorded in multiple media and has produced both structured data (e.g., log files) and unstructured ones (e.g., text, video, images).

In the extracts of three studies, it was demonstrated that there are many issues, relating to analysis of interaction, that necessitate multiple perspectives. Audio recordings of oral communication, video of the whole class or a group of students and observer notes had to be used for interpreting and understanding the fingertip events recorded in the log files. So, analysis tools, like ColAT that interrelate log files and contextual information in these different forms were proved indispensable for supporting and facilitating analysis of activity in these studies.

References

Avouris, N. M., Dimitracopoulou, A., & Komis, V. (2003). On analysis of collaborative problem solving: An object-oriented approach, *Journal of Human Behavior, 19*(2), 147-167.

Avouris, N., Margaritis, M., & Komis, V. (2004). Modelling interaction during small-group synchronous problem-solving activities: The Synergo approach, *2nd Int. Workshop on Designing Computational Models of Collaborative Learning Interaction*, ITS2004, Maceio, Brasil, September 2004. Retrieved June 29 2006, from http://hci.ece.upatras.gr

Avouris, N. , Komis, V. , Margaritis, M. , & Fiotakis, G. (2004). An environment for studying collaborative learning activities. *Journal of International Forum of Educational Technology & Society, 7*(2), 34-41.

Bertelsen, O.W., & Bodker, S. (2003). Activity Theory. In J. M Carroll (Ed.), *HCI Models, Theories and Frameworks*, San Francisco, CA, USA: Morgan Kaufmann.

Cabrera, J. S., Frutos, H. M., Stoica, A. G., Avouris, N., Dimitriadis, Y., Fiotakis, G., & Liveri, K. D. (2005). Mystery in the museum: Collaborative learning activities using handheld devices. *Proc. 7th Int. Conf. on Human Computer Interaction with Mobile Devices & Services*, Salzburg, Austria, September 19 - 22, 2005, vol. 111. ACM Press, New York, NY, 315-318.

Dillenbourg, P., Baker, M., Blaye, M., O' & Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reimann (Eds) *Learning in Humans and Machine: Towards an interdisciplinary learning science*. Oxford: Elsevier. 189-211.

Dix, A., Finlay, J., Abowd, G, & Beale, R. (2004). *Human-Computer Interaction*. 3rd Ed., Prentice Hall.

Edwards, D. & Potter, J. (1992) *Discursive Psychology*. London: Sage.

Fesakis, G., Petrou, A., & Dimitracopoulou, A. (2003). Collaboration activity function: An interaction analysis' tool for Computer-Supported Collaborative Learning activities. In *4th IEEE International Conference on Advanced Learning Technologies (ICALT 2004)*, August 30 - Sept 1, 2004, Joensuu, Finland, 196-200.

Fidas, C., Komis, V., Tzanavaris, S., & Avouris, N. (2005). Heterogeneity of learning material in synchronous computer-supported collaborative modelling, *Computers & Education, 44*(2), 135-154.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Gassner, K., Jansen, M., Harrer, A, Herrmann, K, Hoppe, H. U. (2003). Analysis methods for collaborative models and Activities. In B. Wasson, S. Ludvigsen, U. Hoppe (Eds.), *Designing for Change in Networked Learning Environments*, Proc. CSCL 2003, Dordrecht, Netherlands: Kluwer Academic Pub, 411-420.

Hammersley, M. (1992). *What's wrong with ethnography? Methodological Explorations*. London, Routlege.

A. Harrer, G. Kahrimanis, S. Zeini, L. Bollen, N. Avouris, Is there a way to e-Bologna? Cross-National Collaborative Activities in University Courses, Proceedings *1st European Conference on Technology Enhanced Learning EC-TEL*, Crete, October 1-4, 2006, Lecture Notes in Computer Science, vol. 4227/2006, pp. 140-154, Springer Berlin

Heath, C. (1986). Video analysis: Interactional coordination in movement and speech. *Body Movement and Speech in Medical Interaction*, Cambridge University Press, Cambridge, UK, 1-24.

Heraud, J. M., Marty, J. C., France, L., Carron, T. (2005). Helping. the interpretation of web logs: Application to learning. Scenario improvement. *Proc. Workshop Usage Analysis in Learning Systems*, AIED 2005, Amsterdam, July 2005. Retrieved June 29 2006 from: http://lium-dpuls.iut-laval.univ-lemans.fr/aied-ws/PDFFiles/heraud.pdf

Hulshof, C. D. (2004). Log file analysis. In L.Kimberly Kempf (Ed.), *Encyclopedia of Social Measurement*. Elsevier, 577-583.

Jordan, B. & Henderson, A. (1995). Interaction analysis: Foundations and practice, *Journal of the Learning Sciences, 4*(1), 39-103.

Kahrimanis, G., Papasalouros, A., Avouris, N., & Retalis, S. (2006). A model for interoperability in Computer-Supported Collaborative Learning. *Proc. ICALT 2006 - The 6th IEEE International Conference on Advanced Learning Technologies*, IEEE Publ., July 5-7, 2006 – Kerkrade , Netherlands, 51-55.

Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. *Computer-Supported Collaborative Learning*. Bergen, Norway: Kluwer Academic Publishers. Retrieved June 29 2006 from http://archive-edutice.ccsd.cnrs.fr

Roschelle, J., & Pea, R. (2002). A walk on the WILD side: How wireless handhelds may change computer-supported collaborative learning. *International Journal of Cognition and Technology, 1*(1), 145–168.

Stahl, G. (2001). Rediscovering CSCL. In T. Koschmann, R. Hall, & N. Miyake (Eds.), *CSCL2: Carrying Forward the Conversation*, Hillsdale, NJ, USA: Lawrence Erlbaum Association, 169-181.

Stoica, A., Fiotakis, G., Simarro Cabrera, J., Muñoz Frutos H., Avouris, N., & Dimitriadis, Y. (2005). Usability evaluation of handheld devices: A case study for a museum application. *Proceedings PCI 2005*, Volos, Greece, November 2005, Retrieved June 29 2006 from http://hci.ece.upatras.gr.

Voyiatzaki, E., Christakoudis, C., Margaritis, M., & Avouris, N. (2004). Teaching algorithms in secondary education: A collaborative approach, Proceedings *ED Media 2004, AACE Publ*, Lugano, June 2004, 2781-2789.