Digital Replay system (DRS): A Tool for Interaction Analysis

[*]Patrick Brundell, [○]Paul Tennent, [○]Chris Greenhalgh, [+]Dawn Knight, [○]Andy Crabtree, [*]Claire O'Malley, [*]Shaaron Ainsworth, [*]David Clarke, [+]Ronald Carter and [+]Svenja Adolphs

Email pat.brundell@nottingham.ac.uk

[*]School of Psychology, University of Nottingham, Nottingham, UK, NG7 2RD
[○]School of Computer Science, University of Nottingham, Nottingham, UK, NG7 2RD
[+]School of English Studies, University of Nottingham, Nottingham, UK, NG7 2RD

Abstract

Digital Replay System (DRS) is a software tool being developed by the DReSS node of the UK ESRC-funded National Centre for e-Social Science. It is publically available under an open source license and is designed to support the organisation, synchronised replay, and analysis of complex multimodal corpora including audio, video, dialogue transcripts and system log files. This paper describes the key features of the system with examples from a study of collaborative use of an interactional learning environment.

Introduction

The community of researchers investigating human-human and human-computer interaction collect rich data sets that are becoming ever larger and more diverse as digital recording technologies increase in availability and ease of use. A wide range of both quantitative and qualitative research methodologies are used to investigate these multimodal corpora and software tools are increasingly necessary in order to manage the organisation, replaying, structured and free coding (and annotation) and analysis of these growing data sets. The size and range of data sets produced by varying methodological approaches mean that often the analytical process is tailored to an individual study, and is extremely time consuming. It is also often difficult to relate multiple analyses (for examples see Ainsworth and Burcham, 2007 and Forsyth et al., 2006)

There are a number of tools already in existence to support the analysis of interactions. Examples of such tools appear both commercially and academically. *Transana* (http://www.transana.org) is widely used, and focuses primarily on transcription of both audio and video. Developed by Noldus, *The Observer* (http://www.noldus.com) was designed originally for studying animal behaviour patterns, but has been adopted as a more general coding solution within the social sciences. Mangold International's *INTERACT* (http://www.mangold-international.com) is another observational analysis solution that supports the process of coding videos, then provides some simple visualisation tools to support analysis of the coded data. One final commercial tool supporting real time coding of video is *Studiocode* (http://www.studiocodegroup.com), this has been developed for Apple's

OSX platform. Academic offerings include *I-Observe* (Badre et al, 1995: 101-113), an early project which used a video tape based system and made use of captured event streams to synchronise time-stamped events with the time-code on a video. The ever popular *ANVIL*, developed in 2001 by Michael Kipp at the University of the Saarland was designed as a video annotation tool specifically for the purpose of analysing multimodal corpora (Kipp, 2001: 1367-1370). The *Diver* Project, developed at Stanford University is another tool to support video annotation (Pea et al, 2004: 54-61). Designed to work with a single video, it nevertheless has a unique feature, that of so-called 'dives', where users can manipulate the viewpoint of a video using a virtual camera viewfinder, allowing zooming, panning and rotation of the original video data. Also developed at Stanford University *VACA* provides a toolkit for annotating or coding several simultaneous videos on a timeline representation (Burr, 2006: 622-627). Now in its third version ELAN was developed at the Max-Planck-Institute for Psycholinguistics (Brugman & Russel, 2004: 2065-2068). It is a fairly comprehensive tool for the annotation of video data, primarily in the field of linguistic research. It supports annotation in tiers, which other projects might call tracks, so several simultaneous annotations can be applied to single media object. Current tools tend to be limited in the following ways;

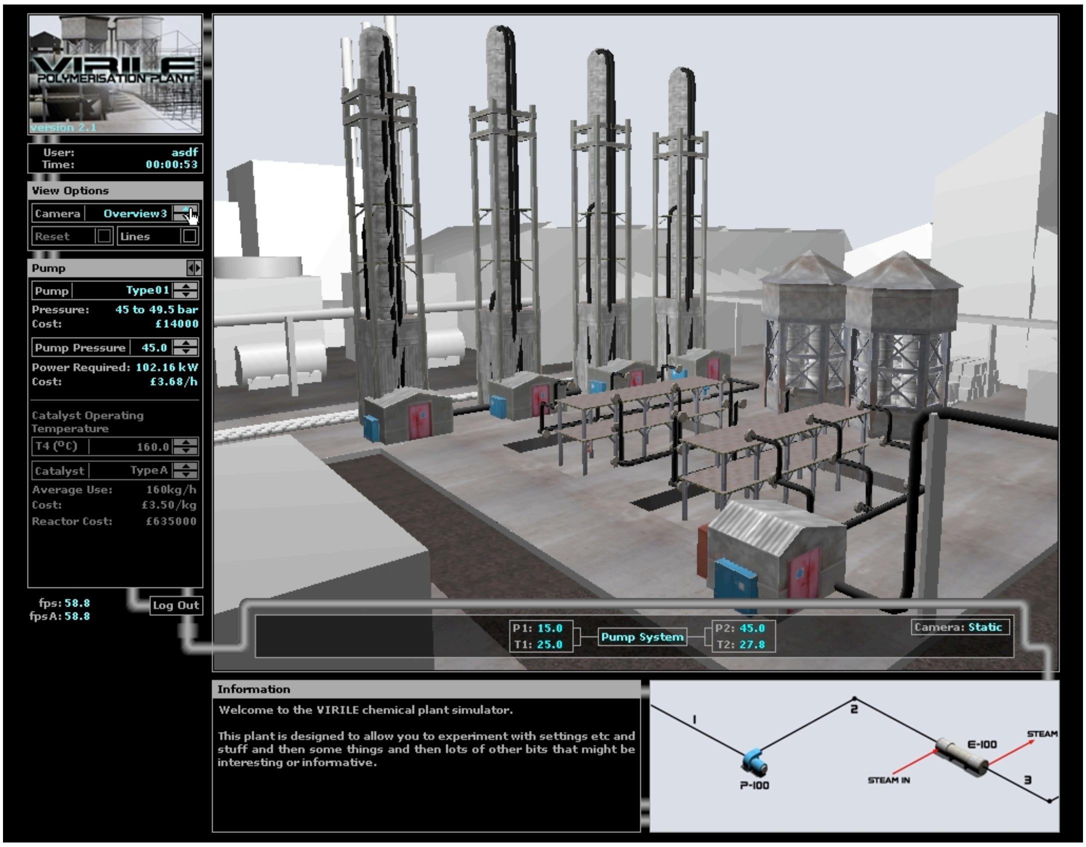
- Limited and/or very specific functionality with a lack of flexibility for a range of user requirements.
- The majority of tools are predominantly home-grown with the associated problems resulting from insufficient time and resources for development such as instability and poor interface design.
- There is often limited support for viewing different data modalities in the tools, specifically, they tend to focus on only on dialogue and/or non-verbal behaviour.
- The majority of tools suffer from a lack of coding flexibility, some tools require programming skills to change existing coding schemes or impose fixed, predetermined coding schemes onto the user.
- Simple coding schemes are generally well supported but complex cross-linking of annotations are generally not possible.
- There is limited potential in most toolkits for the integration and use of automatic coding, speech recognition, spoken language processing, gesture recognition, machine vision or other bolt-on or plug-in functionalities

In this paper we present the novel features of the Digital Replay System (DRS), a tool designed to support the research activities of learning scientists, linguists and ethnographers. We present the underlying ontology and illustrate how DRS can be used to import both raw and structured data, replay synchronised multimodal data and allow further structuring and coding. Finally, we will examine how DRS can be used to create code databases that are flexible and configurable to suit the needs of individual research programs. DRS is in a rapid development cycle and we may describe features which are currently under construction. The functionality of DRS is illustrated with examples of multimodal data analysis from a study of collaborative use of an interactive learning environment for students of chemical engineering.

The VIRILE Study

The study focused on student's experience of using VIRILE: Virtual Reality Polymerisation Plant software. The VIRILE software is a highly realistic and interactive simulation of a large scale industrial chemical processing plant (Schofield, Lester and Wilson, 2004)

Figure 1: VIRILE simulation

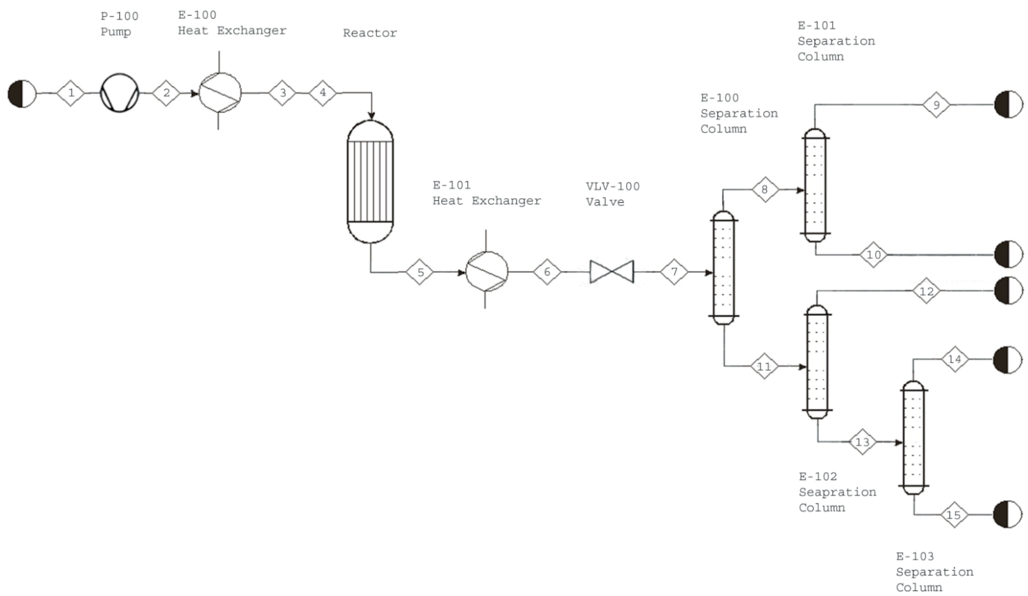


Though the specifics of the study are not relevant in this paper, to give an indication of the scope of data that DRS is intended to be able to assist in organising, synchronising, replaying, annotating and analysing, the study aims were;

To further students' understanding of how large scale industrial plants are designed and to increase their knowledge of what large scale processing components look like; to examine the benefits of exploring a highly realistic and interactive plant simulation to develop students' understanding of how the design of an industrial plant relates to abstract representations of industrial chemical processes (process flow diagrams); to provide students with an rich, engaging and enjoyable learning experience.

The study consisted of three parts. Part one was a written assessment of the chemical engineering knowledge relevant to the functioning of a polymerisation plant, and a pre-test of participant's ability to interpret and produce process flow diagrams and "realistic" drawings of plants. The second part was the learning experience of actively using VIRILE, consisting of two tasks to ensure the participants were able to use the software and had fully explored the visualisation, equipment and processes of the virtual chemical plant. The final task for the students was a test of their ability to interpret and produce process flow diagrams, scale drawings and their more general understanding of the VIRILE chemical plant.

Figure 2: Process flow diagram of the VIRILE plant:



In addition to the tests of learning, recordings were made of students' use of the simulation through system logs and screen recordings. Video and audio recordings were also made of student's interactions with each other.

Digital Replay System (DRS)

DRS consists of a number of essential, integrated functionalities:

- An internal ontology based on the Resource Description Framework (RDF).
- Generalised project organisation including data management and organisation.
- Complex synchronisation of related multimodal data.
- Structured and free-form annotations.
- Support for the import, creation and analysis of textual records.
- Support for the configurable export of structured coding for further analysis.

Internal Data Modelling

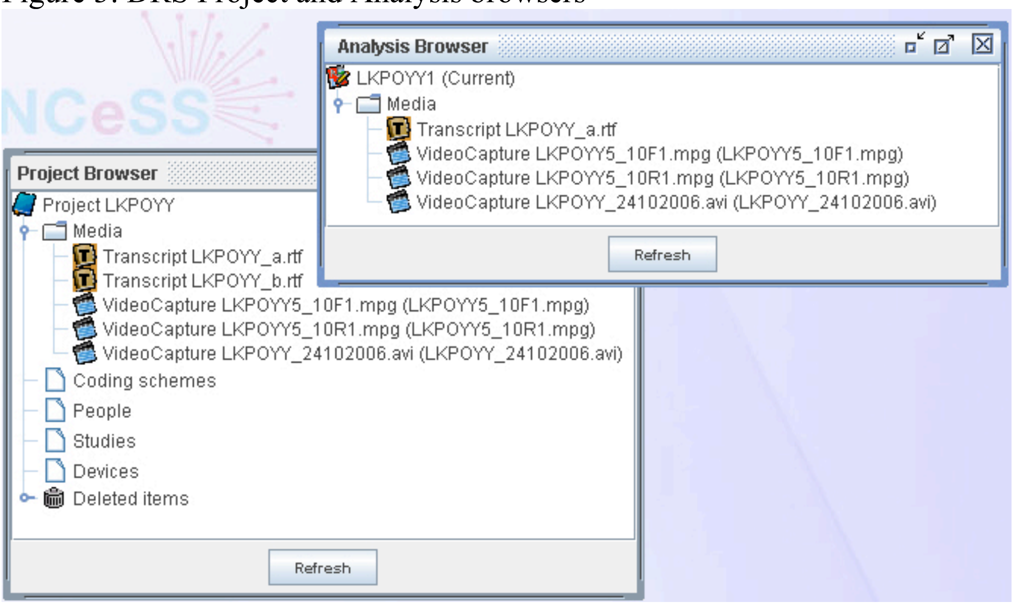
To give it the required flexibility and persistence, DRS uses an extensible and comprehensive data and metadata model based on the W3C's Resource Description Framework (RDF) and Web Ontology Language (OWL) for its ontology. Both are Semantic Web technologies supported by the Open Source JENA RDF library for Java. It includes portions for:

- Its own configuration, e.g., workgroup or standalone, system users and window layouts.
- The files and databases that it is managing.
- Other information that the system depends upon, including projects and analyses, timings and synchronisation of files, coding and annotations.
- Other metadata associated with any of these, for example, recording devices used, participant information etc.

Project Organisation

DRS provides a general project management mechanism whereby multiple files, annotation sets and coding schemes are organised and viewed as distinct "projects" and "analyses". A project might be all the records from a study or a trial. Each project has a project explorer, a simple graphical representation of the project's elements and organisation. Pop-up (context-sensitive) menus give access to secondary functions that are associated with the files (e.g. add to an analysis or export to a new format). A DRS "analysis" (figure 3) is a set of related resources, usually co-temporal and may include audio and video (any format playable by Quicktime), transcripts, annotations, coding schemes and log files.

Figure 3: DRS Project and Analysis browsers



Synchronisation of data

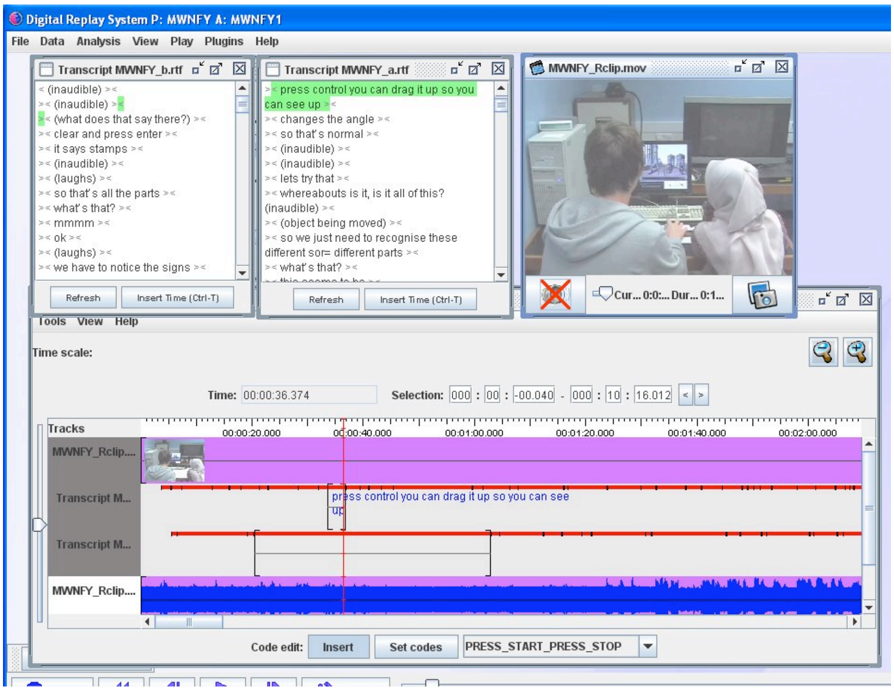
A time-line view is available for each analysis (figure) which gives a visual representation of the temporal extent of media files and associated annotation files, for example coding and transcriptions. The files and annotations (coding and transcriptions) are displayed as tracks. The timeline view is an abstract view of the analysis and within in it, the temporal relationship between media and annotation files may be expressed in a number of ways:

- Media files and analyses may have explicit start dates/times by which their relationship is inferred.
- Media and analyses may have a temporal relationship specified by the analyst and may vary between analyses to express varying perspectives on the same data e.g. to show data from the viewpoint of different participants in non real-time communication.
- Annotation tracks may be explicitly linked to an individual media file, which may be moved independently of any other within the analysis.

Individual tracks are either synchronised by moving them within the analysis time-line or by specifying explicit start times and temporal relations within the synchronisation manager window.

Playback of the analyses is controlled via VCR-like controls allowing normal speed, slow-playback, looping and other such flexible control. Movement within the chronology is available through the time-line cursor and also through annotations files where the analyst may select text/codes and jump to the associated point in the records.

Figure 4: Synchronised transcripts and Media files in DRS



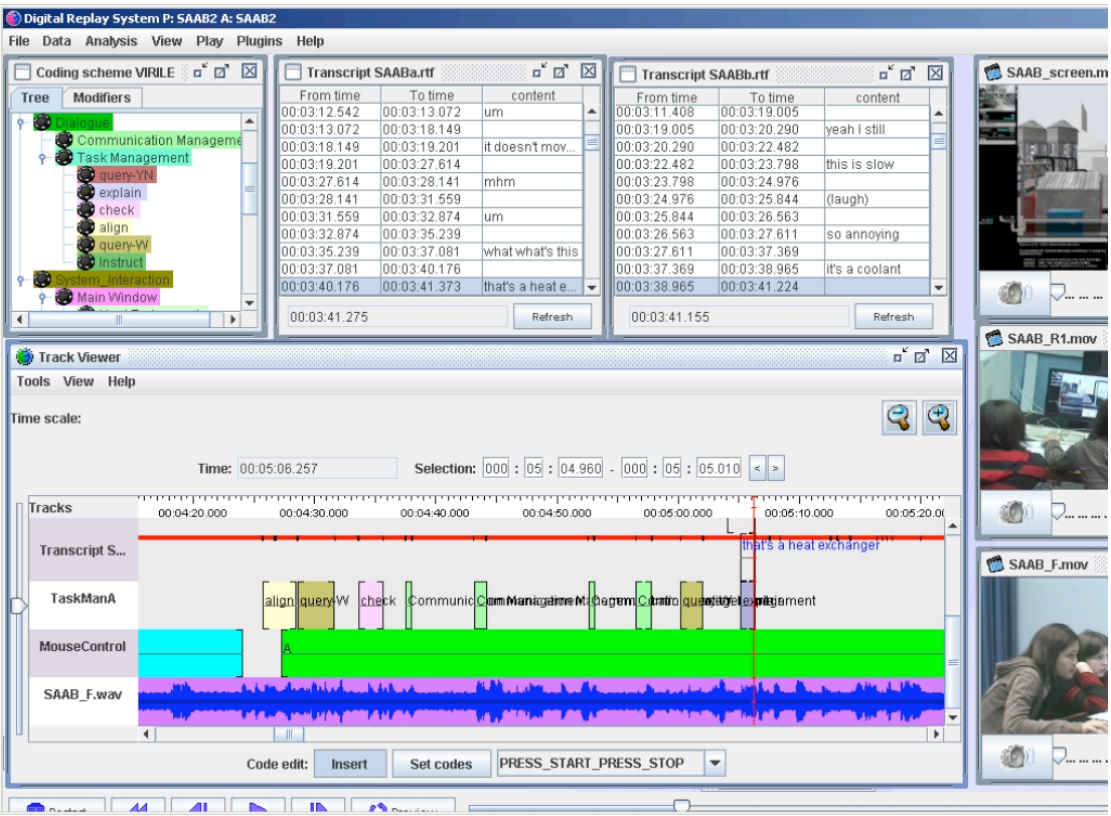
Annotation and coding

The model of annotation defined within the DRS ontology is an extension of that suggested by Bird and Liberman (2001) in which an annotation consists of content associated with a subject. The content of annotations within DRS is text or user defined codes. At present, the main subject of annotations is a region on the analysis timeline, the length of which is defined by the analyst, though in future iterations of DRS there is no reason why other subjects will not be annotatable, defined areas of a still image for example. Related annotations are organised together into annotation sets and appear as a form of media (allowing further analysis as separate entities) within DRS. Annotations are related by their relationship to media within analyses.

Coding

DRS facilitates the structured coding of data. The user typically defines a hierarchical scheme of codes, which are descriptions of behaviour, dialogue etc. (see top left of figure for an example coding scheme used to annotate the VIRILE data. The user then creates a coding track within the track viewer and applies the codes (see the track viewer at the bottom of figure with media, transcription and coding tracks). Codes are defined with a range of possible temporal attributes such as variable duration, fixed duration etc. and are applied to data either during playback or whilst static. The system of code application is flexible and can take place with keyboard shortcuts or by selecting regions of time using the mouse. Code schemes may be modified during the coding process even after coding has begun.

Figure 5: Coded VIRILE data



This functionality supports a multistage iterative process of annotation and coding which might begin with "quick and dirty" exploration of the data where rapid accessing of data and rough qualitative annotation/coding is required in order to identify and note passages of interest and possible variables to be included in a coding scheme and ending with the rigorous application of a finalised coding scheme, structuring data to be used in descriptive and multivariate statistical analyses. This flexibility supports the typical work patterns of interaction analysts where rapid development of an initial coding scheme that is applied to some or all of the data, is subsequently modified and re-applied depending on the descriptive results or visualisations of the data.

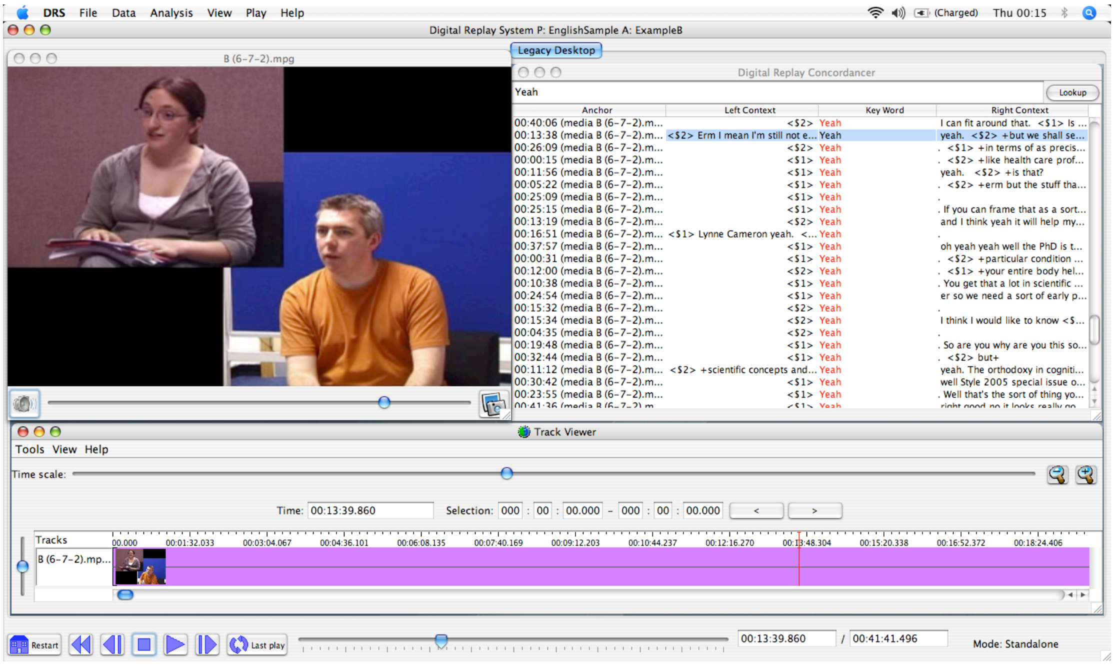
Transcription

Transcriptions are stored internally as a set of non-overlapping free-text annotations and DRS uses time coding information to synchronise them with other media. Time coding may be performed within DRS through a dedicated transcription editor, within the track viewer as a coding track, or it can import files in Rich Text Format (RTF) created in other software such as Transana. Since all annotations are treated in the same manner, it makes the ontology extremely powerful as the tools within DRS will work with any set of annotations, be they codes, free text or transcriptions.

Searching text & annotations

The DRS interface has been integrated with a concordance tool facility, which provides the analyst with the capacity for interrogating coded annotation sets. The concordancer allows searching of a multimodal data set for specific words, phrases or other lexical tags as a 'search term' but also for codes defined within coding schemes. DRS presents a list of occurrences of the search term(s) and their surrounding context within the concordancer window. The analyst may then jump directly to the temporal location of each occurrence within the synchronised record of video and/or audio clips. Currently this functionality is limited to textual transcriptions anchored to media files or individual coding tracks, but in the next software release (in early may) this functionality will be extended to allow searching across all media within an analysis. Since the concordance tool treats textual and coded annotations in the same way internally, they may be presented simultaneously which allows an investigation of both types of annotation in the same manner, applying the same skills and techniques used in the analysis of traditional corpora (for an example see Scott, 1999). This novel multimodal concordancer has led to the need for developing new approaches for coding and tagging language data in order to align textual, video and audio data streams (see Adolphs and Carter, 2007 and Knight, 2006).

Figure 6: The concordance tool in use within the DRS environment



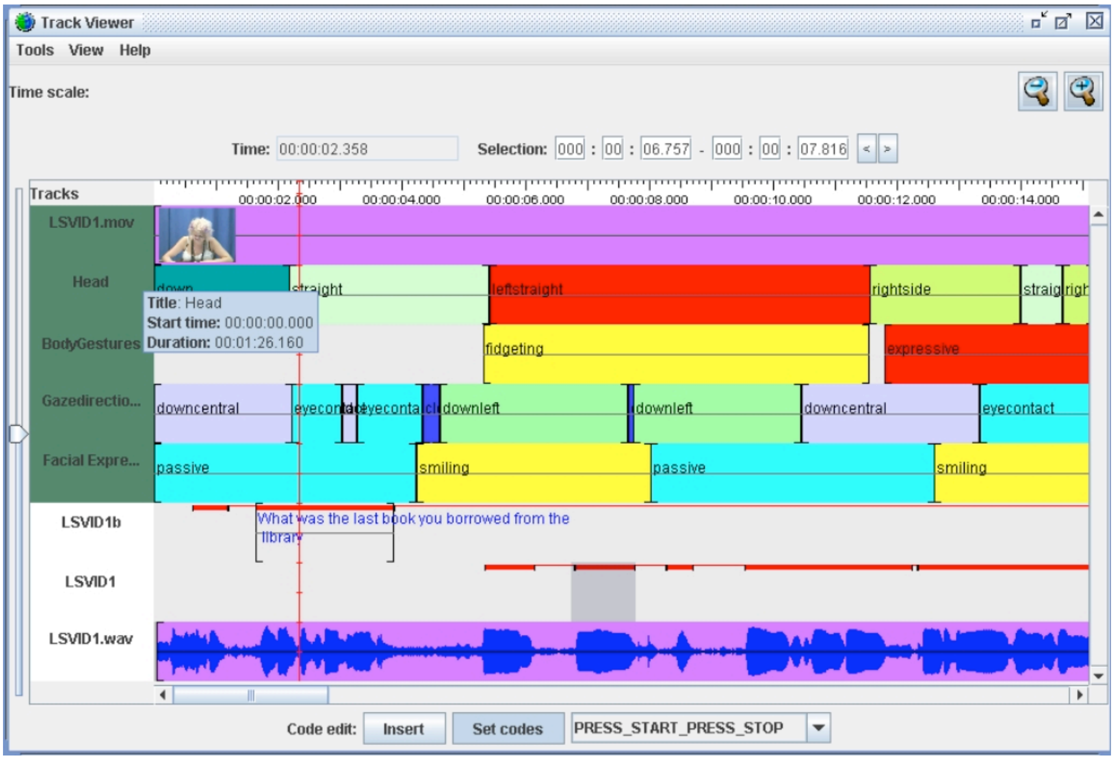
In Figure 6, an analyst has searched for occurrences of the word "yeah" within the transcribed videos of a recorded academic supervision. Each occurrence is highlighted and provided in context within the concordance tool.

Export of structured coding

For many learning science, psychology and interaction researchers, it is the complex interactions between dialogue, non-verbal behaviour and system activity that are of most interest. In the example track viewer screen shot shown below (figure 7), video, audio, transcriptions and coding tracks are shown of a short interaction between two people. The behaviour of one party in the interaction is of particular interest and has been coded for a range of non-verbal behaviours.

The track viewer provides a useful visualisation of the relationships between elements of non-verbal and verbal behaviour, however the ability to export multiple related annotation sets (coding tracks and/or transcripts) is a function that allows quantitative analysis of the data within excel or SPSS. The export processor takes an individual analysis as input and outputs the subject and content of all annotations. Practically, this means that text, code and timing information is outputted into a tabular view window and in a form that allows the use in external statistical analysis packages. The export process has three main phases (all of which the user has varying degrees of control over);

Figure 7: Extensive coding of interactional data



- Segmentation – which determines what each row of the output table will contain and will be dependent on the analysis to be performed with the data. There are two main categories of segmentation type;
    - Fixed period, which divides the analysis into fixed time periods (user defined) and outputs the contents of each coding track and transcription for each time period starting from time 0. This essentially shows what is occurring at regular intervals throughout an analysis.
    - Annotation-based, new rows in the output table are determined by the start or end (or start and end) of annotations. This is effective for showing the times at which events start and end.
- Calculation – which determines the values each row of the output table contains. Columns for each code and transcription are created. There are numerous options for data display here but they fall into two main categories,
    - Text output, the columns are named after the codes and values are the names of the codes applied at the relevant times. Transcription tracks are always output in this manner.
    - Numerical output, the columns are named after the codes tracks and the numerous output values include, "duration", the time within the row's start and end time that the code is applied and "countstart", the number of annotations labelled with this code that start during the time covered by the row.

- Filtration – user defined modification of the output table to omit empty rows or columns

Conclusions and future plans

In this paper we have presented the Digital Replay System, a tool designed to assist research practitioners of CSCW, CSCL. We describe some of the main features including; project, analysis and data management; synchronisation and replay of multimodal data (including audio, video and dialogue transcriptions); structuring, annotation and coding of data, and some of the tools to interrogate raw and structured data. In the workshop we intend to demonstrate these and other features (such as a system log-file handling workbench) with real data from learning and interactional studies. The aim is to garner feedback from experienced researchers and tool developers and expand our understanding of relevant research methodologies. The DReSS team are committed to developing DRS over a period of years and integral to that process is incorporating feedback into future software releases, developing a community of users and providing comprehensive support including the construction of further bespoke plug-ins. These aims marry the software development aspects of DReSS with it's commitments to creating multimodal corpora. The demonstration shall be conducted with reference to existing software solutions.

Digital Replay System is in a rapid development cycle and within the next release due in May 2008 shall be an improved user interface along with refined and extended concordance and multi-track export tools.

Acknowledgments
This work was supported by the ESRC, UK through the grant "Understanding New Forms of Digital Records for e-Social Science" (DReSS node of the NCeSS) and by the EPSRC, UK, through grant EP/C010078/1, "Semantic Media" – Pervasive Annotation for e-Research and EQUATOR IRC, grant GR/N15986/01. Thanks to the other members of the DReSS node for their collaboration and also to staff from the School of Chemical and Environmental Engineering for their help in conducting the VIRILE study.

References

Adolphs, S. and Carter, R. (2007) **Beyond the word: new challenges in analysing corpora of spoken English**, *European Journal of English Studies*, vol. 11 (2), pp. 114-128.

Ainsworth, S., and Burcham, S. (2007). **The impact of text coherence on learning by self-explanation**. *Learning and Instruction*, 17(3), 286-303

Badre, A., Guzdial, M., Hudson, S. and Santos, P. (1995). **A user interface evaluation using synchronized video, visualizations and event trace data**. Software Quality *Journal, 4*(2):101–113.

Brugman, H and Russel, A. (2004). **Annotating multi-media / multi-modal resources with elan**. In *LREC 2004*, pages 2065–2068.

Burr, B. (2006). **Vaca: a tool for qualitative video analysis**. In CHI '06: CHI '06 extended abstracts on *Human factors in computing systems*, pages 622–627, New York, NY, USA, 2006. ACM.

Forsyth, R., Ainsworth, S., Clarke, D., Brundell, P. and O'Malley, C. (2006) **Linguistic-computing methods for analysing digital records of learning**, *Online Proceedings of the 2nd International Conference on e-Social Science*, June 28-30, University of Manchester: ESRC NCeSS.

Kipp, M. (2001) **Anvil - a generic annotation tool for multimodal dialogue**. *In Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pages 1367–1370, Aalborg.

Knight, D. (2006) **Corpora: the next generation**, *Corpus Building and Investigation for the Humanities* (Online Introduction to Corpus Investigative Techniques), AHRC. www.humcorp.bham.ac.uk

Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W., and Hoffert, E. (2004). **The diver project: Interactive digital video repurposing**. *IEEE MultiMedia, 11*(1):54–61, 2004.

Schofield, D., Lester, E. and Wilson, J.A., (2004) **VIRILE: Virtual Reality Interactive Laboratory Experiments**, *Proceedings of Innovation, Good Practice and Research in Engineering Education* Conference, Wolverhampton, 7th – 9th June.

Scott, M. (1999). **Wordsmith Tools**. *Oxford: Oxford University Press*