Analysing video and audio data: existing approaches and new innovations

Elizabeth FitzGerald
Institute of Educational Technology
The Open University
Walton Hall
Milton Keynes
MK7 6AA
+44 (0)1908 659866
e.j.fitzgerald@open.ac.uk

ABSTRACT
Across many subject disciplines, video and audio data are recorded in order to document processes, procedures or interactions. These video and audio data are consequently analysed using a number of techniques, in order to try and make sense of what was happening at the time of the recording, sometimes in relation to initial hypotheses or sometimes in terms of a 'post hoc' analysis where a more grounded approach is used. This paper contains an overview of tools and techniques for examining video data and looks at potential new methods borrowed from the field of learning analytics, related to discourse analysis. Discourse analysis, where conversations and the spoken word are explored and dissected in detail, can provide us with information about the learning context and the ways in which learners interact with people and other resources in their environment.

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human factors, human information processing. H.3.1 [**Content Analysis and Indexing**]: Linguistic processing. K.3.1 [**Computer Uses in Education**]: Collaborative learning; Computer-assisted instruction; Computer-managed instruction.

General Terms
Measurement, Documentation, Experimentation, Human Factors, Languages, Theory.

Keywords
Video analysis; audio analysis; discourse analysis; qualitative evaluation.

1. INTRODUCTION
Collection of video data is often used when studying how people interact with each other and with technology or artefacts. It can provide research communities with a powerful way to collect, share and analyse complex processes of human interaction. This short paper seeks to provide a brief overview of how video collection and analysis may be used in educational technology settings, particularly looking at human-human and human-computer interactions. It also explores more novel and innovative ways in which video (and particularly audio) data can be analysed, with a view to presenting this at the workshop and exploring its potential with example video clips. Throughout this paper, the term 'video' is referred to frequently; however it should be remembered that this also includes a substantial *audio* component, which can either be analysed as a separate entity or in conjunction with other aspects of the video (e.g. stills or segments of 'whole' video).

2. RECORDING, STORING AND SHARING VIDEO
A plethora of devices exist to record video data easily and cheaply, from dedicated video cameras (at varying degrees of cost), to those available on common handheld devices e.g. smartphones or other mobile phones, or digital cameras. For example, head-mounted cameras can be used in fieldwork settings [2] or for capturing surgical procedures (e.g. the Deixis project: http://www.siumed.edu/call/html/deixis.html). The Diver project (Digital Interactive Video Exploration & Reflection, http://diver.stanford.edu) used a set of 5 cameras to collect a 360-degree record of activity.

In addition, a number of websites and dedicated smartphone 'apps' have emerged in recent years, to support the creation, hosting and sharing of videos (e.g. YouTube; Tumblr etc). Many of these are in relation to the 'social media' phenomenon, where users create and share multimedia for professional or personal/leisure use. Whilst these resources are not the focus of this paper, they serve to illustrate how ubiquitous the idea of video, as a publicly-created and share communication medium, has become. Moreover, it shows how easy it is to create and share video using multiple devices and hence also through multiple surfaces. Many people are familiar with the idea of recording and sharing video; it is hoped that that this increased familiarity amongst the general populace leads to greater acceptance and more naturalistic settings, where video is recorded by researchers and used in studies where those subjects in the video are the main points of interest. In any case, video diaries and video-recorded observations/user trials have a long-standing tradition in several disciplinary research areas and so it is hoped that this paper would

Surface Learning Workshop 2012, March 18-20, 2012, Bristol, UK.

be of interest to a large number of academics and researchers who record and analyse video data in their everyday work.

3. ANALYSING VIDEO DATA: EXISTING TOOLS AND PRACTICES

Video captured *in situ* can contain a great richness of information, often revealing subtle yet important incidents relating to the interactions between people and technology. A number of tools have been developed to assist human analysis of recorded video.

For example, Transana (http://www.transana.org) is a popular software package designed to facilitate the transcription and qualitative analysis of video and audio data. It allows collaborative analysis between academics working from different locations, through sharing of analytic markup and access to shared video.

The Diver project (mentioned previously in this paper) provides a suite of web-based exploration and annotation tools, which also enables several researchers working in collaboration, to make selections and share their ongoing analyses of the video data [12].

Another example of a tool used for video analysis is the 'Video Traces' project (http://faculty.washington.edu/reedstev/vt.html), that allowed people to create layers of voice and pointing/tracing on top of existing video recordings. These traces could then be shared with other colleagues, or with those featured in the video and then used as prompts for reflection [14].

DRS (Digital Replay System) enabled the synchronization, replay, and analysis of audio and video recordings [4]. DRS also allowed these to be combined with system logs, which recorded interaction within computational environments. SMS messages, interaction in virtual environments, GPS data or data from body sensors, for example, could be imported into DRS, synchronized with conventional recordings, and be replayed alongside them.

As well as specific tools to enable video analysis, it is also important to consider how to go about performing the analysis itself. A number of techniques can be used to examine video data and it largely depends on what line of inquiry the researcher wishes to take. For example, Erickson [7] suggests three sets of alternative guidelines:

1. Whole-to-part inductive approach
2. Part-to-whole inductive approach
3. Manifest content approach

The first approach is suggested for identifying patterns in the data where there are no initial hypotheses, theories or predictions, thus employing a more grounded method. The second approach is in direct contrast to this, where the video data is scrutinized for specific types of events and is most relevant where the research is driven by existing questions, theories or hypotheses about those events. The manifest content approach is where interactions are selected and examined, that focus around particular subject or pedagogical content.

Another way in which video can be analysed is through the use of the Critical Incident Technique [9]. This technique was originally used in aviation, requiring pilots to record incidents in the cockpit. These incidents were then analysed and conclusions were drawn, in order to produce design principles for future methods for training new pilots. Most pertinent to this paper is the use of the technique in HCI and education. In HCI, the focus of the incident is often on events where something either goes unexpectedly well, or badly [5]. These incidents can then be used to inform the design of further iterations of the software in question. In education, the technique has been adapted to uncover breakthroughs and breakdowns in teaching and learning activity which are then probed through retrospective interviews with the participants [1]. Critical incident analysis has also been used as a way in which teachers can engage in reflective practice through analysis of videotaped lessons, as shown by Brantley-Dias *et al* [3], who worked with science teachers to promote 'reflection-in-action' [13] and 'reflection-for-action' [10].

4. POTENTIAL NEW INNOVATIONS IN VIDEO AND AUDIO ANALYSIS

Learning analytics is a relatively new but fast-moving research field, aimed at analysing the vast amounts of digital data created in relation to learners and their activities, particularly when considering their interactions with the Internet and associated systems (VLEs, databases containing student information etc.). Learning analytics tends to seek *qualitative* understanding of the context and meaning of such information, in contrast to mere web analytics, that often present more *quantitative* data [8].

Recent work in learner discourse analytics shows some promising techniques that could be applied to video data analysis. From a sociocultural perspective, it can be argued that the quality of dialogue taking place between learners can have a direct impact on their level of educational success or failure [11]. In addition, sensemaking is said to be intrinsically tied to its social context [15]. Mercer *et al* mentions three social modes of thinking, used by groups of face-to-face learners: cumulative, disputational and exploratory talk. Exploratory talk is considered the most advantageous of the three, as it enables learners to develop shared understanding through reasoned discussions, challenging ideas and examining/evaluating evidence. Disputational talk, as its name suggests, contains high levels of disagreement whilst cumulative dialogue centres around the addition of contributions of others, without much challenge or criticism. Initial work by Ferguson and Buckingham Shum showed that synchronous text chat could be manually analysed by using exploratory dialogue analysis, indicating periods of meaningful discussion between participants [8].

De Liddo *et al* go one step further and have documented how *automatic* text analysis can be carried out, using discourse-centric machine annotation to detect sentences that include "summarizing" functions or "contrasting ideas". Sentences with "contrasting ideas" can be further categorized into sub-classes such as novelty, surprise, emerging issue, open question or importance [6].

It is hoped that these novel approaches from the field of learning analytics might be applied to the analysis of video/audio transcripts. If shown to have some practical use, taking into account the degree of accuracy, this could be an exciting first step towards automated transcript analysis, showing at which points, and between which learners, effective and meaningful dialogue occurs.

5. SUMMARY AND CONCLUSIONS

This paper has made an initial attempt to bring together ideas and practices behind the creation and analysis of video and audio data, as a qualitative method to inform research into technology-

enhanced learning (TEL) and human-computer interaction. It has made a brief foray into existing tools and techniques for examining video data and has also proposed new ways in which video/audio data may be analysed. Whilst this paper is still somewhat skeletal in nature, it is hoped that it will form the basis of further interesting discussions and potential new techniques to be trialled in the future, that will in time provide much more in-depth work into video/audio analysis that will be of vital importance to researchers across many disciplines and particularly to those working in HCI and TEL.

6. ACKNOWLEDGMENTS

Many thanks to Rebecca Ferguson and Simon Buckingham-Shum, both from the Open University, for sharing their recent work and insights into learning analytics. Also thanks to Brian Elliston, a PhD student in the Horizon Doctoral Training Centre at the University of Nottingham, for helping me originally expand my horizons into discourse analysis and the notion of 'talk-in-interaction'.

7. REFERENCES

[1] Anastopoulou, S., Sharples, M., Wright, M., Martin, H., Ainsworth, S., Benford, S., Crook, C., Greenhalgh, C. and O'Malley, C., Learning 21st Century Science in Context with Mobile Technologies. *Proceedings of the mLearn 2008 Conference: The bridge from text to context*, (Wolverhampton, UK, 2008), 12-19, 2008.

[2] Beddall-Hill, N.L. Witnessing learning in mobile settings using a head mounted camera. In E. Brown, editor, *Education in the wild: contextual and location-based mobile learning in action*. University of Nottingham: Learning Sciences Research Institute (LSRI), Nottingham, UK, 2011, 39-42.

[3] Brantley-Dias, L., Dias, M., Frisch, J.K. and Rushton, G., The Role of Digital Video and Critical Incident Analysis In Learning to Teach Science. *Proceedings of the American Educational Research Association Annual Meeting*, (New York City, New York, 2008), 2008.

[4] Brundell, P., Tennent, P., Greenhalgh, C., Knight, D., Crabtree, A., O'Malley, C., Ainsworth, S., Clarke, D., Carter, R. and Adolphs, S., Digital Replay System (DRS) - a tool for interaction analysis. *Proceedings of the 2008 International Conference on Learning Sciences (Workshop on Interaction Analysis)*, (Utrecht, 2008), 2008.

[5] Carroll, J.M., Koenemann-Belliveau, J., Rosson, M.B. and Singley, M.K., Critical incidents and critical themes in empirical usability evaluation. *Proceedings of the HCI-93 Conference: People and Computers VIII (British Computer Society Conference Series)*, (1993), Cambridge, U.K.: Cambridge University Press, 279-292, 1993.

[6] De Liddo, A., Sándor, Á. and Buckingham Shum, S. Contested Collective Intelligence: rationale, technologies, and a human-machine annotation study. *Computer Supported Cooperative Work (CSCW) (in press)*, 2012.

[7] Erickson, F. Definition and analysis of data from videotape: Some research procedures and their rationales. In J.L. Green, G. Camilli and P.B. Elmore, editors, *Handbook of complementary methods in education research*. Erlbaum, Mahwah, NJ, 2006, 177-205.

[8] Ferguson, R. and Buckingham Shum, S., Learning analytics to identify exploratory dialogue within synchronous text chat. *Proceedings of the 1st International Conference on Learning Analytics & Knowledge (LAK 2011)*, (Banff, Alberta, 2011), 2011.

[9] Flannigan, J.C. The critical incident technique. *Psychological Bulletin*, 51 (4), 327-358, 1954.

[10] Killion, J.P. and Todnem, G.R. Capturing complexity: A typology of reflective practice for teacher education. *Teaching and Teacher Education*, 18, 73-85, 2002.

[11] Mercer, N. Sociocultural discourse analysis: analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, 1 (2), 137-168, 2004.

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

[13] Schön, D.A. *The Reflective Practitioner: How Professionals Think in Action*. Maurice Temple Smith, London, 1983.

[14] Stevens, R., Cherry, G. and Fournier, J., VideoTraces: rich media annotations for learning and teaching. *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community (CSCL '02)*, (Boulder, Colorado, 2002), International Society of the Learning Sciences, 525-526, 2002.

[15] Weick, K.E. *Sensemaking in Organizations*. Sage Publications, Thousand Oaks, CA., 1995.