

# Video Annotation – Improving Assessment of Transient Educational Events

*Matthew Butler, Tim Zapart, and Raymond Li  
Monash University, Melbourne, Australia*

[matthew.butler@infotech.monash.edu.au](mailto:matthew.butler@infotech.monash.edu.au)  
[tim.zapart@infotech.monash.edu.au](mailto:tim.zapart@infotech.monash.edu.au) [ray.li08@bigpond.net.au](mailto:ray.li08@bigpond.net.au)

## Abstract

Annotation of video content has been commonplace in the entertainment industry for many years and is now becoming a valuable tool within the business world. Unfortunately its use in education has to date been limited. Although research and development is being undertaken to apply video annotation techniques to assessment and both software and hardware exists to facilitate this process, it must be acknowledged that these solutions are generally cost prohibitive for educational use.

This paper will investigate simple video annotation methods for assessment of transient state events in the education context such as presentations. The authors will introduce a number of existing uses of video annotation, as well as discuss the educational context this can be placed within, highlighting fundamental concerns with some assessment practices. A framework for a solution involving video annotation techniques will be discussed, along with practical demonstration of a prototyped solution and discussion of further application.

**Keywords:** video annotation, multimedia, transient events, Macromedia Director

## Introduction

The nature of multimedia, as both a discipline and as content, has grown much more complex over the last 5 years. No longer does it simply represent a combination of sound and video, but now also encompasses information rich applications. Part of the reason for this growth is that multimedia now truly represents its title, applications that span both creative and technical depth.

One industry that has been at the cutting edge of multimedia development is undoubtedly the “entertainment” industry. Whether it be in areas such as filmmaking, gaming, web, or television, the entertainment industry has always been an important leader and developer. One important area of development in recent times has been that of creating information rich multimedia, namely the annotation of video and audio content, the focus of this paper.

---

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact [Publisher@InformingScience.org](mailto:Publisher@InformingScience.org) to request redistribution permission.

Video annotation is the process of making time based notes corresponding to elements of the video footage, “usually a way to describe the content for later retrieval” (Correia & Chambel, pg. 151). Schroeter, Hunter, and Kosovic (2003) describe it as “subjective comments, notes, explanations or external remarks that can be attached to a document or a selected part of a document without actually modifying the document” (pg. 1).

Both definitions highlight it as a form of metadata annotation. A primary use to date has been in the television industry, where live footage is often annotated so that “highlights” sequences can be quickly cut together. The business world is beginning to realise the power of this ability, with a number of commercial packages now available for capturing and annotating important business events, such as meetings and seminars, for ease of later review and improving business performance.

Although much research is being undertaken into annotating video content for many different purposes, mainly archival interests and professional development, an application that has been explored very little to date is within the domain of educational assessment. This is surprising as educators are constantly looking for alternate assessment items and techniques, which has resulted in heavy use of events such as presentations, interviews and seminar participation as primary assessment items. The fundamental problem with these items, just as in the business context, is that they are transient events, existing only once unless captured by video or audio means. The process of re-evaluation of the items can be a long and laborious process, however Dowman, Tablan, Ursu, Cunningham, and Popov (2005) point out that annotation affords the opportunity to access and review material that would otherwise be difficult due to the lack of description available.

This paper will investigate the use of simple video annotation methods for transient state events for the use of assessment in the education context. The authors will begin by introducing a number of existing uses of video annotation, providing a brief introduction to the technology and its current use. The educational context will then be discussed, highlighting fundamental concerns with some assessment practices and the need for a solution involving video annotation techniques. A framework will then be developed and discussed, along with practical demonstration of a prototyped solution and discussion of further application.

### **Existing Uses of Video Annotation**

As mentioned in the previous section, the use of video annotation is already well ingrained in certain industries. The entertainment industry has been using different manners of annotation for many years, primarily in the world of television. Broadcasters generally annotate material that will be used later for either immediate “highlights” purposes, or for archiving. Assfalg, Bertini, Colombo, Del Bimbo, and Nunziati (2003) describe two approaches to this annotation process: “Production Logging” in which producers will mark up an event live, to note shots to be edited into highlights packages and “Posterity Logging” in which librarians make detailed annotation of video tape for long term reuse, where depth and historical context is also noted.

The level of detail of notes that are being made to video footage are increasing in complexity and indeed many researchers are investigating ways that this annotation process can embed this greater amount of metadata. Similarly, researchers are also investigating ways in which the process of annotation can be streamlined, allowing this increased detail to be able to be made in real-time and even collaboratively. Researchers, such as Schroeter et. al. (2003), are obtaining significant funding in order to develop Collaborative Video Indexing and Annotation systems.

However it is not only academic researchers who are investigating the possibilities of improved video annotation. Indeed many companies realise the benefit of being able to provide simple systems to the average consumer to use for this purpose. In fact many offer both software and hardware to facilitate this process. Hardware such as the Advanced Micro Peripherals VAC104+ (“VAC104+”, n. d.) allows a user to accept a video stream and generate computer graphics and notation to blend with the video.

Similarly, software exists that allows a user to annotate a live video feed. The sporting environment has made extensive use of this, as exemplified by video analysis software offered by *DataPro-*

ject (“Data Project”, n. d.). This use is now becoming increasingly popular in the business environment, to document video footage of meetings and even job interviews. For example, systems such as *Studiocode* (“SportsTec”, n. d.) are being marketed directly to businesses for use in this manner.

It cannot be long before the average television viewer begins performing annotation of their own. Simon Perry, publisher of *Digital Lifestyles* magazine is championing this convergence: “Individuals creating their own metadata around a broadcast TV programme. I think it is a fascinating area... When people are able to, through a broadcast, drop metadata in whether they are able to put markers in a time line... it becomes very interesting. That is another thing that is great about convergence” (Lovatt, n.d).

## The Educational Context

The assessment of students in the university environment is under constant revision and scrutiny. Although exams have been a staple of traditional student assessment, educational theorists are constantly extolling the virtues of alternate assessment items. Issues of *formative* versus *summative* assessment and *deep* versus *surface* learning, are commonly raised in discussion of the outdated role of exams and the need for varied assessment. Biggs (1999) defines formative assessment as focusing on feedback that “operates both to improve the learning of individual students and to improve teaching” (p. 142). Unlike exams, the focus is clearly on assessment items that provide feedback to students and guidance in their learning, rather than simply providing the examiner a tool by which to grade or rank students (i.e. summative assessment). Similarly, exams often assess a surface understanding of material. By the very nature of a time-limited, paper-based assessment medium, a deep, analytical understanding of material cannot often be demonstrated.

Industry bodies also are playing a major role in the changing nature of assessment. Future employers of university graduates are playing a large part in dictating the skill set required by students on graduation and rather than a sole focus on industry specific skills and knowledge, broader aptitudes such as presentation skills, interpersonal skills and project management, are being demanded. Assessment tools such as exams do not provide the appropriate ability to demonstrate and assess these skills, again facilitating a need to look outside traditional assessment techniques.

As a result of these factors, presentations, seminars and interviews have now become commonplace in university assessment. As Crème highlights in “Assessment for Learning in Higher Education” (Knight, 1995), “It (a presentation assessment) encourages deep learning. Having to prepare a seminar makes us think as thoroughly about the material as if we were doing an essay on it in order to think of questions, important issues, how to structure a learning session” (p. 138). There is however inherent difficulties and fundamental issues in assessment items of this nature.

To begin, assessment items such as presentations, seminar participation and the like are transient in nature, meaning that unlike an exam or an essay they exist once and then are no longer kept by the examiner. Obviously these events can be recorded, however in many cases this is not done and even if it is, recorded versions can be difficult to reassess (especially longer seminars) given the time needed, environmental factors that are not captured, or generally elements that are difficult to reassess.

Adding further difficulty is that assessments of these items often have wildly varying criteria that need to be assessed concurrently. A presentation for example requires not only that the content is accurate, but that the “presentation” element is sound. So at one particular moment, an examiner may have to be assessing verbal content, visual content such as overhead aids, personal mannerisms of the presenter and audience reaction, and whilst assessing these elements, must be re-

Recording all relevant notes for not only assessment purpose, but also student feedback. If a strict criterion-referenced assessment model is being used (as dictated by sound assessment practice), then the process becomes virtually impossible. How does an examiner assess all elements mentioned above, while trying to relate it back to a set of predetermined criteria, in a way that is meaningful and accurate? Assessment and feedback quality can only suffer as a result.

## Solution Framework and Design

As indicated in the discussion of current use of annotation, solutions already exist both in research circles and commercially, that can begin to solve this problem. Both software and hardware exists that can facilitate this process. However it must be acknowledged that for the most part, these solutions are cost prohibitive for educational use.

Multimedia development tools now exist that can combine video with database activity. Applications such as Macromedia Director and Macromedia Flash are designed to build multi-modal applications, that combine video, audio and user interaction. Considering that the educational assessment application of video annotation does not have any of the demands that may be placed on more sophisticated uses of annotation, such as in the entertainment industry, these tools may provide a perfectly good platform for developing an application for annotation of student work. The sophisticated hardware and software solutions begin to prove to be both too complex and too costly for consideration, when a cheaper and easier solution may be at hand.

A completely real-time system could be created to handle annotation of student presentation work using these development applications, however need not be at this stage. In most cases, student assessors will be part of a presentation audience, therefore will be viewing the event in person. As a result, a basic system need only to record time-based assessment of key criteria while the event is happening, which can then be aligned with video of the event at a later stage.

Therefore, in order to demonstrate the use of video annotation for assessment purposes, a simple system has been developed. As indicated, the system is fairly rudimentary, being a prototype to evaluate the worth of a video annotation system rather than yet provide all possible functionality and convenience.

The basic logical process for the video annotation prototype is as follows:

- Set up a new assessment annotation session
- Capture the event using a standard digital video camera
- During the event record time stamped criterion-based assessment and ad hoc comments in the application (with use of laptop)
- Record all data to an application data store
- Digitise and embed video footage to the annotations
- Review video and annotations on demand

Based on this process, the developed system prototype was designed to set up a new session, record and store all annotation data, allow for embedding of video footage and facilitate later review of both video and annotations.

As the embedding of video was a major component, along with the view to ultimately have real-time streaming of video into the main application interface, Macromedia Director was chosen as a development tool. Consideration was given to a number of other development environments, however Director was chosen due to a number of factors including its video handling capabilities, data handling capabilities, ease of multi-platform development and developer experience.

Interface design for the application was kept relatively simple to provide ease of assessment. Wizards were developed for session setup in order to guide the user through the process. This process allows the system to not only capture all required preliminary information, but also to apply set assessment criteria to the new session, pre-determined by the user and applied based on the event being assessed, be it a presentation, interview, or seminar session. This can ensure a greater quality of assessment during the evaluation process by providing the assessor with a strict criterion-based assessment framework. A single interface screen was designed to capture all video annotation. This was to reduce distraction and confusion during the already hectic assessment process. Examples of this and other interfaces can be seen in the following section.

## Demonstration of Application

As discussed, the developed prototype is relatively simple in design, however it successfully demonstrates the desired features and meets the initial goals of development. In order to analyse the prototype, it is easiest to run through the basic system functionality.

The screenshot shows the 'timeStampAssessment' application interface. At the top, there is a navigation bar with 'new session', 'review session', and 'quit' options. Below this, the text 'Create new session wizard: Page 1/3' is displayed. The main content area features a 'Create new session' window with a clock icon. The window contains the following fields and options:

- Session ID:** A text input field containing the value '50'.
- Date:** A text input field containing the value '10/11/05'.
- Staff Name:** A list box containing the names: Butler, Hegman, Smith, Sofra, Hopman, and Zapart.
- Unit Code:** An empty text input field.
- next:** A button located at the bottom right of the window.

Figure 1 – Session Creation

Figure 1 shows an initial screen for establishing a new video annotation session and is the first screen of a multi-stage “wizard” for doing so. As the underlying purpose of the system is to use it for assessment of student presentations, seminars and the like, the key information captured in this process includes time and date, staff members assessing the session, unit details, student participants, session type (presentation, interview, seminar) and the like. For the purpose of the prototype, very little information is actually required, however a full system should not have any more complex data requirements in the establishment of a new video annotation session.

Once the session has been established, the assessor then has the option to begin the session. Figure 2 shows the main assessment screen and is the key interface during the assessment process. As detailed in the previous section, the interface provides the assessor with facility for both criterion-based assessment and ad hoc comments. The criterion-based assessment ensures rigidity of the assessment process and allows the assessor to easily award marks to each element. The ad hoc session allows the assessor to provide detailed feedback if need be, importantly time stamped for

## Video Annotation

reference against the transient event. Again, note at this time that video is not yet being fed directly into the application.

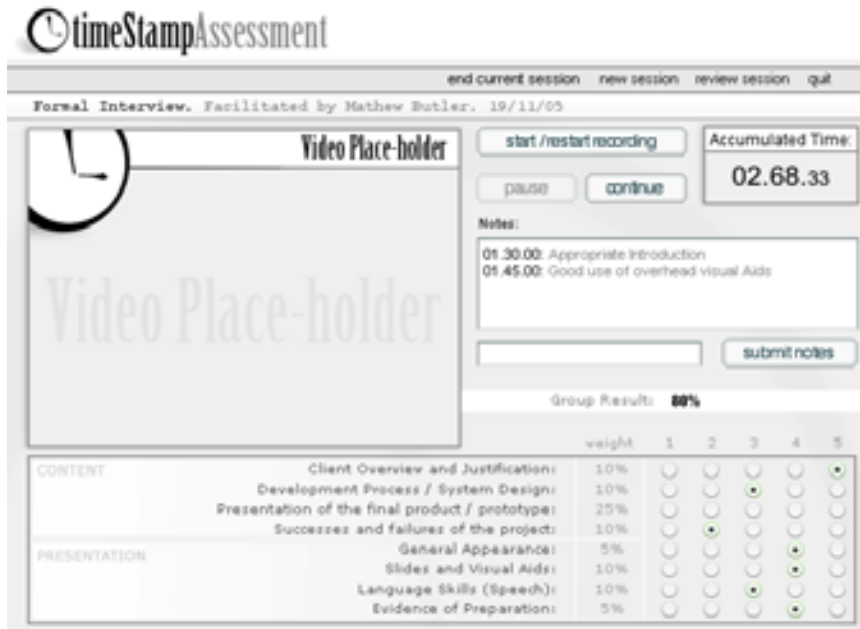


Figure 2 – Assessment



Figure 3 – Review Session

Once the session has concluded, the assessor can end the session, in which final data is written to the datastore. It is at this stage that the user may insert the actual video captured into the application. At this stage of the prototype, this is a manual process, with video footage needing to be

digitised and then linked into the interface and explicit session. The authors acknowledge this as an immediate limitation of the prototype, however for demonstration purposes this does not compromise assessment functionality.

A fundamentally important element of the prototype is the ability to review previously recorded sessions. Figure 3 shows the screen to search for a session by a number of different criteria. Once the session is found, then the review interface is simply the same as the primary assessment interface, as shown in Figure 2. All prior assessment and notes can be seen, along with the now integrated video. As discussed in previous sections, this improves quality of assessment for transient events such as presentations by allowing reassessment of the event by either the same academic or other assessors. This feature also has the ability to provide very explicit feedback to students. Post-event feedback for presentations and interviews can be extremely difficult to learn from, as immediately after the event feedback is removed from its specific context. Post-session review allows the student to reflect on their performance, along with the feedback given and provide a stimulus for further discussion and analysis.

## Conclusion / Further Research

The developed prototype is successful in demonstrating the basic advantages of using a video annotation process for the assessment of transient events such as presentations, interviews and seminars. It improves the assessment of these events by:

- Capturing the transient events for later review
- Capturing time-stamped assessment of the event, allowing for second examination, student reflection and more detailed feedback
- Making the process of assessment less complex

Further research would obviously require rigid assessment of the prototype, to examine to what extent both quality of assessment and feedback is improved. The authors also intend to make the video annotation system completely real-time, streaming and capturing video directly in the application, eliminating a simple yet time-consuming part of the current process. Knowledge of the Director development environment suggests that this is not a complex task.

Further development also intends to allow for the annotation of other events, rather than restricting the use to basic assessment items such as presentations and interviews. Annotating audio, for example subject lectures, could be of great benefit, allowing students to make more meaningful notes for later review.

## References

- Assfalg, J., Bertini, M., Colombo, C. Del Bimbo, A. & Nunziati, W (2003). Automatic interpretation of soccer video for highlights extraction and annotation. *Proceedings of SAC 2003*, Melbourne Florida, USA, pp. 769 – 773
- Biggs, J. (1999). *Teaching for quality learning at university* (1<sup>st</sup> ed.). Buckingham Society for Research into Higher Education, Open University Press
- Correia, N & Chambel, T. (1999). Active video watching using annotation. *ACM Multimedia '99* (part 2), pp. 151 – 154
- Data Project. (n. d.). Retrieved 30/11/2005 from <http://www.dataproject.com>
- Dowman, M., Tablan, V., Ursu, C., Cunningham, H. & Popov, B. (2005). Semantically enhanced television news through web and video integration. *Second European Semantic Web Conference (ESWC'05)*, Crete, 2005

Knight, P. (Ed.). (1995). *Assessment for learning in higher education*. London: Kogan Page.

Lovatt, F. (n. d.) The IBC digital lifestyles interviews – Simon Perry. Retrieved 20/09/2005 from <http://digital-lifestyles.info>

Schroeter, R., Hunter, J. & Kosovic, D. (2003). Vannotea – A collaborative video for indexing, annotation and discussion system for broadband networks. *K-CAP 2003 Workshop on Knowledge Markup and Semantic Annotation*, Florida, 2003.

SportsTec (n. d). Retrieved 20/09/2005 from <http://www.sportstecinternational.com>

VAC104+ - PC104+ Video Annotation Controller (n. d.). Retrieved 20/09/2005 from <http://ampltd.com/prod/vac104p.html>

## Biographies

**Matthew Butler** is a lecturer and PhD student within the Berwick School of Information Technology, at Monash University, Melbourne Australia. His teaching areas include introductory programming, project management and information technology management. His interests in research revolve primarily around innovative teaching and assessment practices, with his PhD topic examining methods of improving off campus delivery and understanding of key introductory programming concepts.

**Tim Zapart** is a lecturer at the Berwick School of Information Technology, at Monash University, Australia. He has a Bachelor of Multimedia and a Master of Information Technology. His research interests are centered around the use of Machinima (real-time animation scripted within an electronic game engine using game-based production methods) to improve visualisation and reduce the time and cost of developing scenarios for integration within e-training applications. Tim also teaches in the areas of multimedia production and digital media authoring.

Currently **Raymond Li** is the director of e-promote Pty Ltd, a Melbourne based SME specialized in e-training and project management consultancy. Early in 2006, he retired from Monash University where he was a Senior Lecturer. His research interests include Experience Design Game-based e-learning, Web based collaborative Network Diagram creation including Automatic Graph Layout using Genetic Algorithm, Automatic Essay Scoring using Latent Semantic Analysis, Knowledge Based Lesson Learned system for Project Management, Mobile Agent and Web Services, 3D Visualization of Simulations using Cellular Automata. Beside teaching and research, he had successfully completed many industrial based e-learning development projects. His clients include the Victoria Police, Bristol-Myers Squibb and Southern Health in Melbourne, Australia.