LearnCanada
Multimedia Software Workpackage

October 2002

Bruno Emond & Martin Brooks, NRC

bruno.emond@nrc.ca
martin.brooks@nrc.ca

www.learncanada.com

Table of Contents

1. Overview

The LearnCanada Multimedia Software Workpackage was hosted at the NRC Institute for Information Technology as part of the Broadband Visual Communication Research Program. The Workpackage objective was to research broadband multimedia applications in support of teacher professional development. Work was carried out in close collaboration with educators whenever possible; they defined application requirements and evaluated a series of prototypes. Teacher response is documented in the LearnCanada evaluator's report.

Multimedia software requirements evolved during LearnCanada, as a consequence of both the evolution of teachers' experience and the rapid development of streamed media technology and standards. This important constraint on software development was anticipated at project start, and led to a rapid prototyping approach using low-cost commercial-off-the-shelf (COTS), open source and original software. In the latter case, Common Lisp, a programming environment especially well-suited to prototyping was chosen. Multimedia tools and content developed in LearnCanada were focused on server-based multimedia document composition and delivery using World Wide Web Consortium[1] standards.

The following multimedia applications were developed in LearnCanada. The remainder of this document describes each of these applications in detail and lists publications, presentations, and multimedia content created during the project.

Private Video Server (PVS)[2]

Videoconference spans distance, but requires scheduling – a difficulty that increases nonlinearly with the number of sites and participants. "Private Video" entails the notions that video is captured and produced at near-zero cost, and is viewed only by trusted colleagues and mentors. PVS includes functionality for content management, video delivery, and dynamic attachment of text annotation to the video, to streamline communication regarding the content. Research included automatic summarization of video speech content, for use in skimming, browsing, and searching video content.

PVS can:

- allow teachers to asynchronously[3] share, discuss and reflect on behaviours captured in videos of themselves and their class.
- facilitate distance collaboration, by providing visibility of activities at the individual sites in the intervals between synchronous sessions.
- improve human performance by supporting timely expert feedback.
- bind virtual communities by providing a common point of reference on community activity.

Videoconference Extension through Webcast (VEW)

A webcast stream containing all videoconference sites allows additional people to observe. LearnCanada's prototype webcast solution, "Videoconference Extension through Webcast" (VEW), includes functionality allowing all those observing to communicate with one another, i.e. to discuss the contents of the videoconference. By having a person at a videoconference site monitor the webcast, communication back to the videoconference is established. In effect, this creates a penumbra of partially participating people around the focal area of fully participating people, thus greatly increasing videoconference reach.

Videoconference Capture

Video records of videoconferences are an important source for PVS, analysis and communication. Video capture and presentation in a manner that effectively communicates behaviours and content in broadband learning sessions is a challenging problem.

LearnCanada's multimedia software intellectual property is owned by NRC, and is available to the research community via a cost-free license.

2. Private Video Server (PVS)

Note: Earlier LearnCanada documentation referred to PVS as "Video Sharing and Annotation (VAS)".

The idea behind the Private Video Server originated in discussion with educators at a LearnCanada project proposal planning meeting. The designated purpose was as a reflective tool, allowing teachers to share their classroom experiences with trusted peers and mentors, who would provide constructive criticism and encouragement. The requirements expressed at that time were:

- near-zero cost of video capture and production;
- privacy vis a vis access to the content;
- tools facilitating discussion around the content.

The following sections describe two PVS prototypes developed in light of authentic teacher requirements, a description of LearnCanada research into automatic metadata generation using speech understanding, followed by a discussion of LearnCanada's experience and future directions for asynchronous tools.

2.1 PVS Version 1

Following project start, a series of meetings with LearnCanada educators, focused on PVS, indicated

[1] http://www.w3.org/

[2] PVS was called "Video Annotation and Sharing" (VAS) in earlier LearnCanada documentation.

[3] "Asynchronous" means that group access is not simultaneous. Examples of asynchronous communication include email and voice mail.

that schools could:

- capture, produce, render, upload, and set access permissions for their video segments;
- download over CA*net 3, view videos, and add text annotation.

The first version, PVS-1, was prototyped and released to the teachers for field testing in February 2001. At that time, an expert-led videoconference training session for all school boards was held, including breakout sessions where teachers downloaded and annotated video, accessing the experts on-demand.

2.1.1 User Functionality

PVS-1 was intended as a reflective tool. A document "Broadband Visual Communication for the Reflective Practitioner" was made available to the teachers. The following paragraphs are excerpted here:

Level one:

The teacher practitioner video-captures her/his teaching behaviour and reviews it alone. The teacher annotates personal comments using a project based leaning rubric as a guide. The video files are securely stored in a personal mailbox for future comparative analysis. The files can be used to develop a digital personal teaching portfolio, and to allow the teacher to reflect on her/his teaching.

Level two:

The teacher practitioner captures her/his teaching behaviour and reviews it alone, and subsequently with students. They reflect on teacher and student behaviour in the classroom using a project based learning rubric to guide their discussions. These discussions could also be captured for the teacher to review and engage in additional reflection on their collaborative evaluation of each other. The video files are securely stored in a personal mailbox for future comparative analysis and content for their teaching portfolios.

Level three:

Two teacher practitioners capture their teaching behaviour and send it to each other for review and feedback. Each annotates comments using a project based learning rubric. Their video files are securely stored in their personal mailboxes for future comparative analysis and content for their teaching portfolios. The teachers sign an agreement that only they are allowed to review the other's video files.

Level four:

Practitioner teams distributed among multiple sites capture their teaching behaviour and send it to each other for review and feedback. Each annotates the other team members' video with feedback comments using a project based learning rubric as a guide and send each other their respective analysis. They schedule a follow up broadband encounter to share their combine learning experiences and build their knowledge of project based learning together. Selected files of their individual teaching could be shared and streamed to all sites for discussion and collaborative analysis. The team's combined rubric assessment of selected files could also be projected to evaluate any differences in their individual analysis and evaluation. These team encounters can also be capture for future reflection and analysis either by individual team members or by the team as a whole. The team's video files are securely stored for future comparative analysis and content for their team's collaborative teaching portfolio. The teachers sign an agreement that only they are allowed to review each other's video files.

Level five:

A mentor is invited to support levels one to four. The mentor acts as a reflective coach, facilitates authentic assessment and suggests areas for future exploration and development of innovative teaching behaviour. The mentor signs an agreement that only he/she and the teachers are allowed to review the video files.

Level six:

An educational researcher or research team is invited to support levels one to five. Practitioners collaborate with researchers in the development of advanced pedagogy and tools to support a broadband enabled learning environment. Practitioners' user-requirements drive the research, development and evaluation process. The teachers sign an informed consent form which outlines recognized standards for ethical research on human subjects.

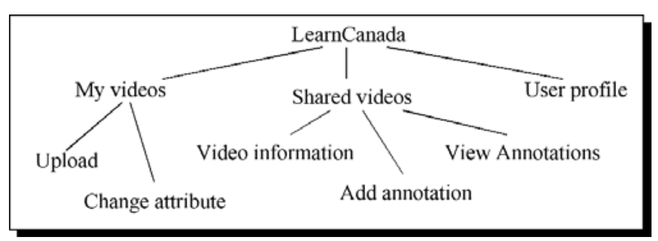

Figure 1. Overview of PVS functionality.

Figure 1 presents a general overview of user functionality for PVS-1. Figures 2 through 6 show web pages generated by VSA.

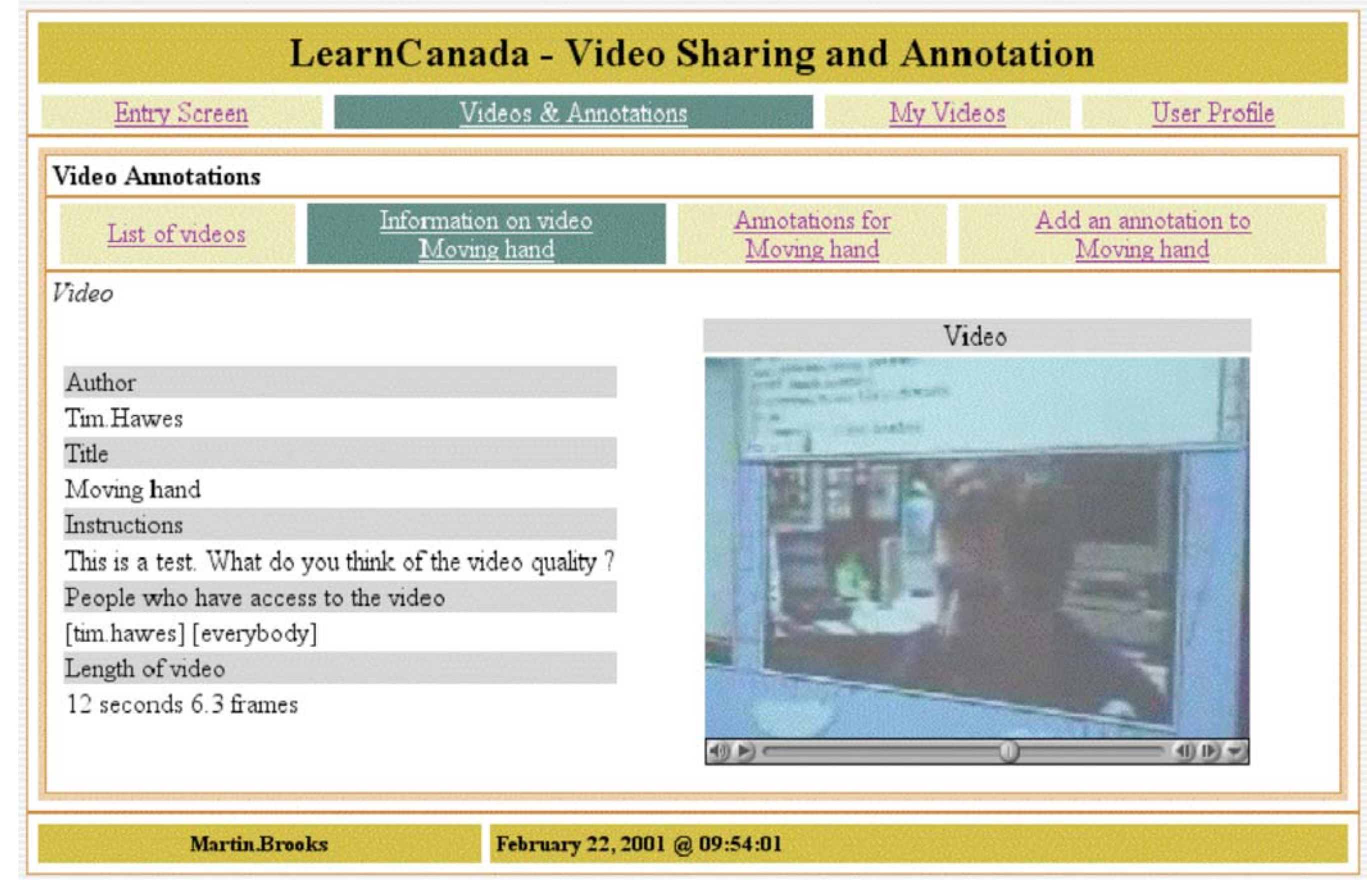

Figure 2: Viewing a video.

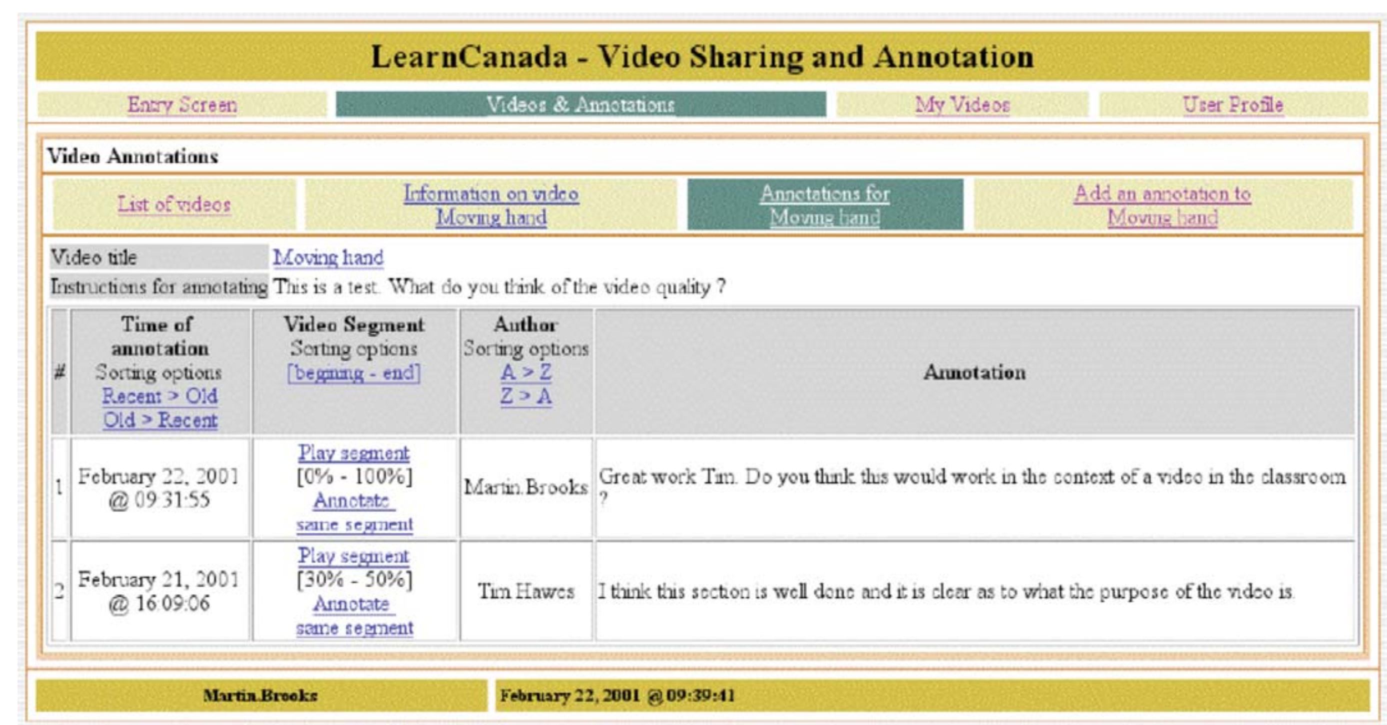

Figure 3. Annotations provided by people who viewed the video.

Figure 4. Adding an annotation.

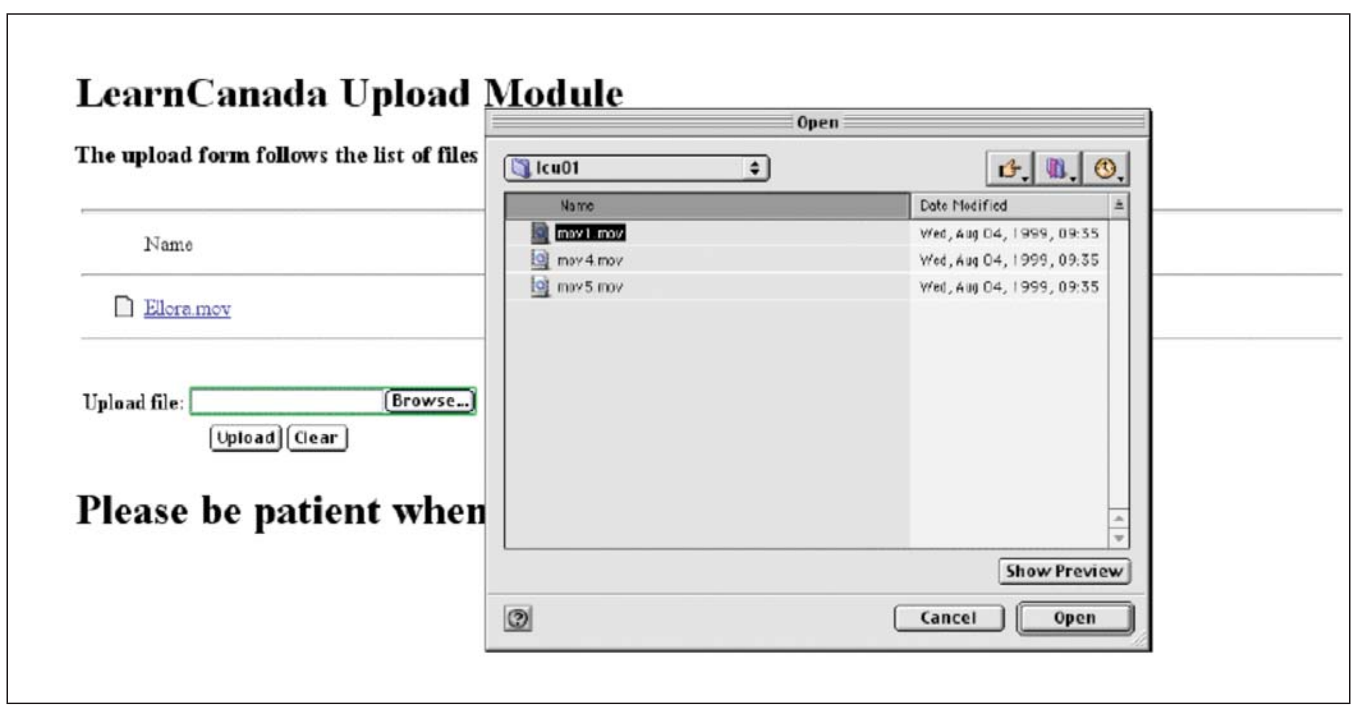

Figure 5. Selecting a video to upload to the server.

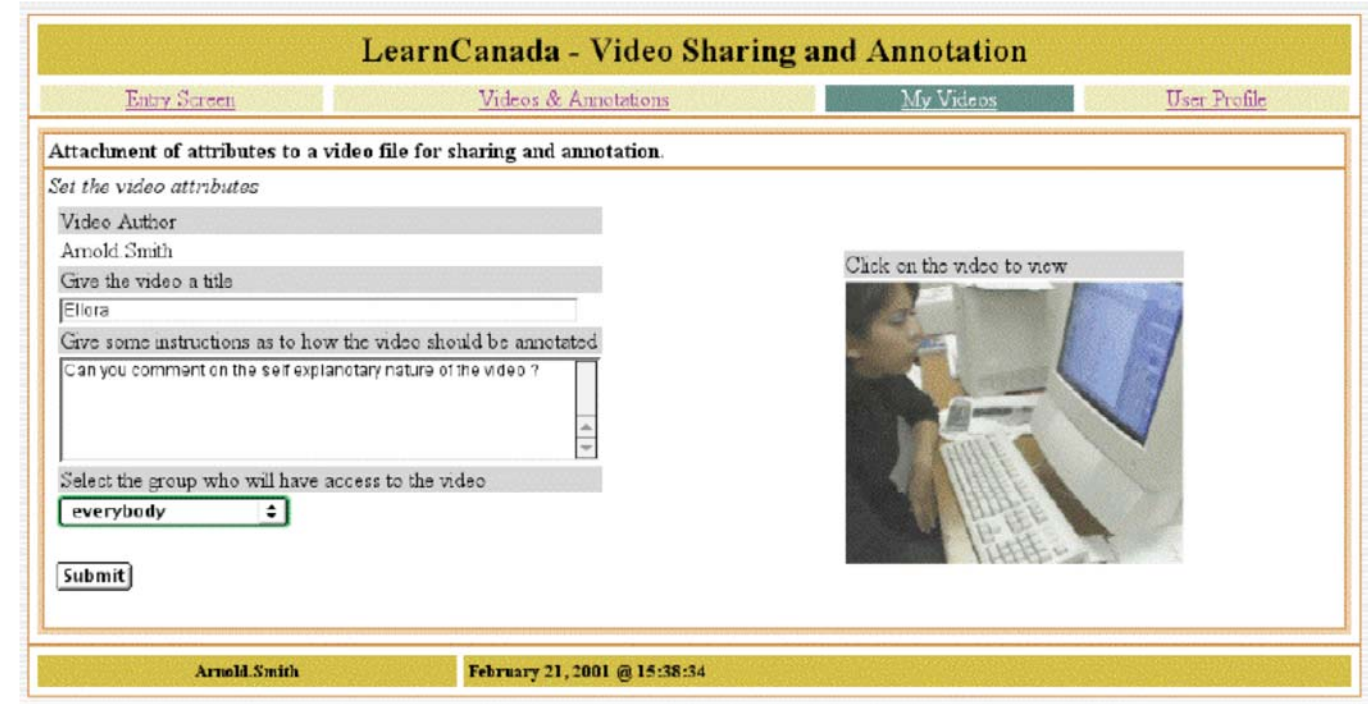

Figure 6. Changing the properties of a video for sharing.

2.1.2 Technology

The PVS-1 server was built using the technologies, standards and commercial-off-the-shelf software (COTS) in Table 1. PVS functionality included video upload by ftp, video sharing by downloading video to a web-page plug-in, and textual video annotation via a forms-based web interface.

Platform	Network	Software	Standards
• 400MHz Macintosh G3 • MacOS 9.1	NCIT*net[4] provided 100 Mb to CA*net 3	• cl-http[5] • WebStar (ftp)[6] • original code in Macintosh Common Lisp[7]	• HTML 4.0[8] • QuickTime[9]

Table 1: PVS-1 technology components

The PVS client is a generic web browser with a QuickTime plugin, available at no cost for Windows and Macintosh workstations.

Hosting PVS on NCIT-net was a consequence of NRC's 7Mb bandwidth connection to CA*net 3. NCIT-net not only provided 100Mb access via CA*net 3, but also access via commodity Internet.

QuickTime was chosen as the multimedia format due to the low cost of rendering software for Windows and Macintosh platforms. Alternatives, such as Windows Media and Real Player, were much more expensive. Since every school needed to render their own video, cost was an important consideration.

The cl-http open source web server dynamically generates HTML for each url request. This was appropriate for PVS, since pages change as annotations are added.

Original software was implemented using the Common Lisp Object System, part of Macintosh Common Lisp. Information about the video content and the associated annotations were stored in an object-oriented database.

2.1.3 Field trials

PVS-1 was provided to teachers during spring 2001. Following the initial training session, on-demand support was provided to teachers, and comments regarding usability and functionality were collected.

---

[4] **www.ncit.ca**

[5] Common Lisp Hypermedia Server – **www.ai.mit.edu/projects/iiip/doc/cl-http/home-page.html.** Open source from MIT Artificial Intelligence Lab, written in Common Lisp.

[6] **www.webstar.com**

[7] **www.digitool.com**

[8] HyperText Markup Language – **www.w3.org/TR/html4/**

[9] **www.apple.com/quicktime**

A variety of constructive suggestions concerning media viewing and the web application were provided. However, one issue stood out above all others: Most of the schools were not able to arrange for a Windows or Macintosh workstation on CA*net 3. The workstation that was connected to CA*net 3 ran Linux, for use as Isabel videoconferencing. There is no QuickTime client for Linux. Consequently, teachers were not able to download videos over CA*net 3. Instead, they downloaded over commodity Internet, including both high speed (cable or DSL) and 56k dialup, both from school and home. Long download times hindered practical application of PVS-1, but valuable feedback was obtained, nevertheless.

2.2 PVS Version 2

Summer 2001 was used to solve the video download problem encountered in PVS-1. Version 2 (PVS-2) replaced video download with video streaming. Each video was made available at three different streaming rates, appropriate for 56k dialup, cable/DSL, and CA*net 3 access. Selection of appropriate streaming rate was provided automatically. Additionally, support for the Synchronized Media Integration Language (SMIL) was included as a precursor to new client interfaces that could potentially both increase annotation richness and simplify use.

2.2.1 User Functionality

PVS-2 usage differed from PVS-1's only with respect to video upload. Because of the complexity of creating and configuring three different renderings of each video, the ftp upload module was removed. Rendering then became a central service, wherein schools would send digital video tapes to NRC.

2.2.2 Technology

The PVS-2 server was built using the technologies, standards and commercial-off-the-shelf software (COTS) shown in Table 2. PVS functionality included video sharing by streaming at one of three automatically determined rates to a webpage plug-in, and textual video annotation via a forms-based web interface.

Hardware	Network	Software	Standards
• 400MHz Macintosh G3 • MacOS 9.1 • 400MHz Macintosh G4 • MacOS 10.0	NCIT-net NRC LAN	• cl-http • original code in Macintosh Common Lisp • QuickTime Streaming Server[11]	• HTML 4.0 • SMIL 1.0[10] • QuickTime

Table 2: PVS-2 technology components

The change to streamed delivery had the following consequences for the PVS-2 server:

- PVS-2 was composed of two servers, each on a separate machine: a video streaming server (QuickTime Streaming Server running under MacOS 10.0), and a web server implemented in cl-http (Common Lisp Hypermedia Server) running under MacOS 9.1.
- PVS-2 utilized the QuickTime "reference movie" feature, resulting in automatic selection of the bandwidth-appropriate stream. This feature requires that client QuickTime Player preferences indicate connectivity bandwidth.
- The new implementation supported the Synchronized Multimedia Integration Language (SMIL 1.0) as a means to structure both video viewing and video annotation.

Inclusion of SMIL was not reflected in user functionality, but holds significant potential for future functionality. SMIL offers control over the presentation of multimedia elements that go well beyond the capability of the html imbed tag.

SMIL 1.0 is a World Wide Web Consortium standard supported by QuickTime Player and RealPlayer. SMIL provides synchronization of multimedia, e.g. nested sequential and simultaneous video displays, and also provides conditional display through user interaction via standard interface tools implemented directly in the media player. This would allow teachers to add annotation in the forms of text, video, documents, and web resources, a feature teachers had requested.

PVS-2 can deliver the streamed media reference as an embedded movie or SMIL file into a webpage, or deliver the movie/SMIL file directly to a media player. Addition of annotations was restricted to the web delivery mode. Figure 7 shows the flow of streamed media between QuickTime Streaming Server (QTSS), PVS-2, and client applications.

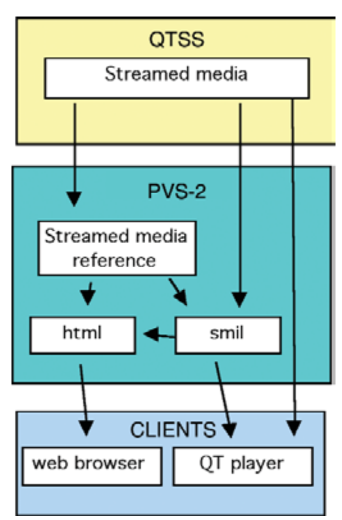

Figure 7: Flow of streamed media is PVS-2.

[10] **www.w3.org/AudioVideo/**

[11] **www.apple.com/quicktime/products/qtss**

2.2.3 Field trials

The plan for Fall 2001 was extensive testing of the new streaming environment with teachers in three schools under various bandwidth conditions. At that time, the LearnCanada teachers' professional development work was largely centred on designing and implementing student projects. Many of these projects were multi-school, using Isabel videoconference to coordinate the streams of collaboration at their respective schools. Together with the teachers, it was determined that at this time PVS-2's most immediate impact would be as a collaboration tool, providing visibility of activities at the individual sites in the intervals between synchronous sessions.

Teacher field trials were cut short by network technical problems. NCIT-net function became unreliable starting at the end of October 2001. Consequently, PVS-2 trials were continued by two university students hired specifically for this purpose. Results indicated that streaming eliminated the usability issues that plagued PVS-1, and that many smaller teacher-suggested improvements had been successfully achieved.

2.3 Automatic generation of metatdata for PVS

Metadata supports media management, and enables functionality for searching and browsing media collections. Metadata includes information about the media's provenance, history, and content. Future technology maturity will make search-by-content a powerful aid to utilization of video for learning.

LearnCanada carried out research into automatic generation of video content metadata, using speech recognition to generate partial transcriptions to which a text summarization tool could be applied. Results are promising.

The research began with a review of the three existing products which include speech tools. Products from Virage, MediaSite, and Excalibur were evaluated through vendor presentations. Each cost in the $100-150k range, and are usually purchased by media-asset companies, such as television broadcasters. These systems offer similar functionality: Extracted speech forms the index of a search engine; typing a one-word query results in a set of video segments where that word has been recognized as having been spoken.

Large-vocabulary speech understanding achieves high accuracy only under controlled audio conditions after having been trained on the speaker's voice. The evaluated systems met their minimum speech reliability specifications only on video meeting broadcast quality standards; further improvement was possible through training. Since LearnCanada's Private Video is not broadcast quality, we decided to eschew the expensive products and focus on creating a research prototype.

A speech engine was needed for the prototype. IBM ViaVoice and Dragon Naturally Speaking were readily available. Visits to three speech technology companies – VoiceIQ, LocusDialog and SpeechWorks – provided insight into more advanced engines. However, practical considerations led us to work with Dragon Naturally Speaking. Extensive experience with this had already been gained in speech application research at NRC's Interactive Information Group (at the Institute for Information Technology), with whom we were collaborating.

In particular, a research prototype application called "Extractor for Speech" had been developed, based on application of the Dragon speech engine to non-production-quality video speech, together with a mature product of NRC research called Extractor.[12] Extractor scans a text document and returns a handful of keyphrases, i.e. one-to-three word phrases appearing in the text which most summarize what the text is about. These phrases mimic the data typically found in the author-provided keyphrase field in many document metadata schemes.

Extractor for Speech uses the speech engine to create a partial transcription of the video's soundtrack. Despite relatively low word recognition rates, Extractor is able to identify meaningful summarizing keyphrases in the transcription. Details can be found in the publication "Extracting keyphrases from spoken audio documents"[13].

Extractor for Speech produces a file containing keyphrases extracted from the sound track as well as a time stamp reflecting their scope and position. A PVS module was developed to produce a set of SMIL and HTML files from a movie's keyphrase file, thus providing keyphrase-based browsing through the movie. Figure 8 shows one interface, giving each keyphrase's occurrence and a timeline representing the speech segment for which the keyphrase is a summary. Clicking on the timeline plays the associated video segment. Figure 9 shows a different interface, where a movie has been split into segments of equal length, and keyphrases associated with each segment are displayed. These developments have not yet been field-trialed.

[12] **www.dbi-tech.com/dbi_extractor.asp**

[13] **See last section of this report.**

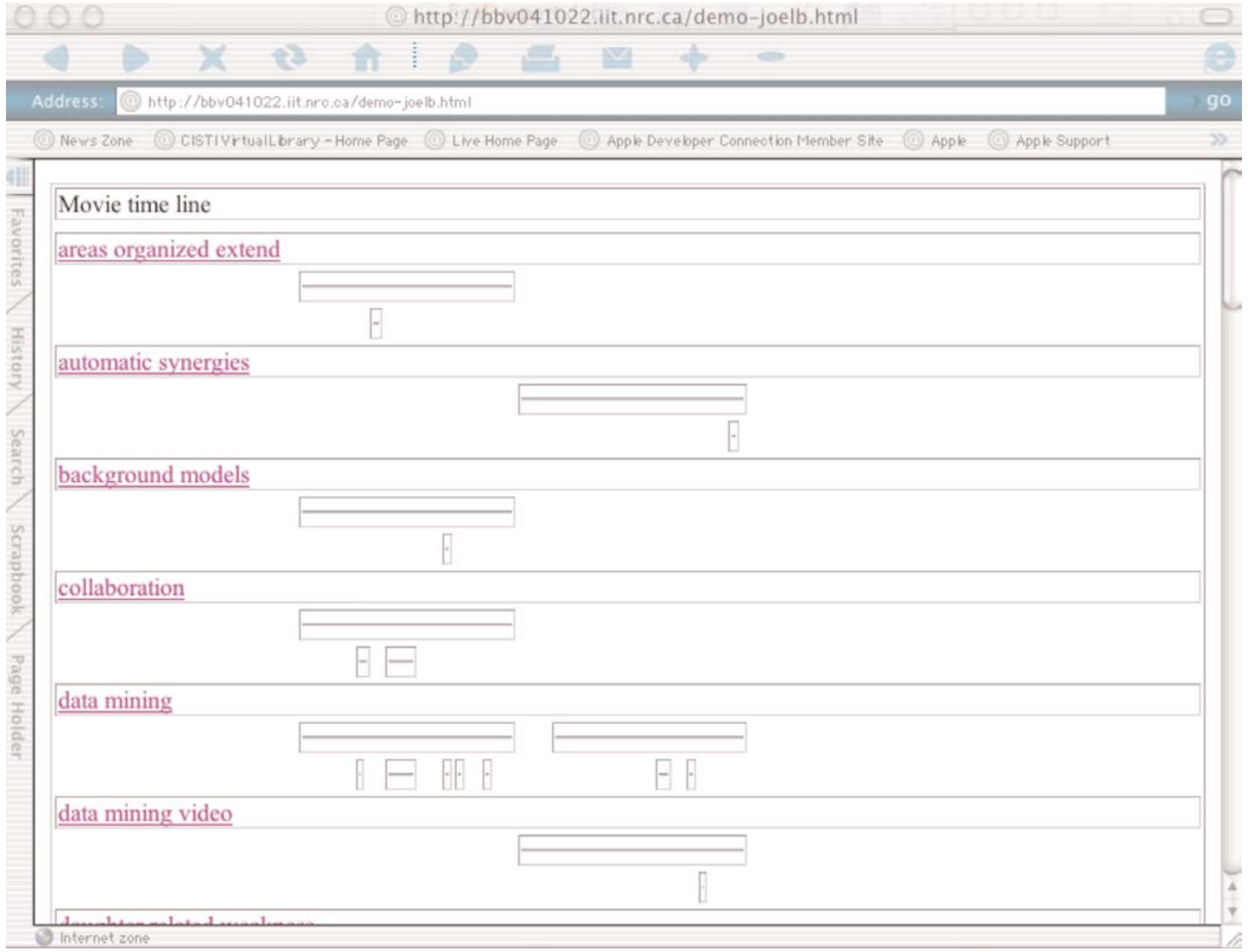

Figure 8: Browsing by keyphrase scope and occurrence.

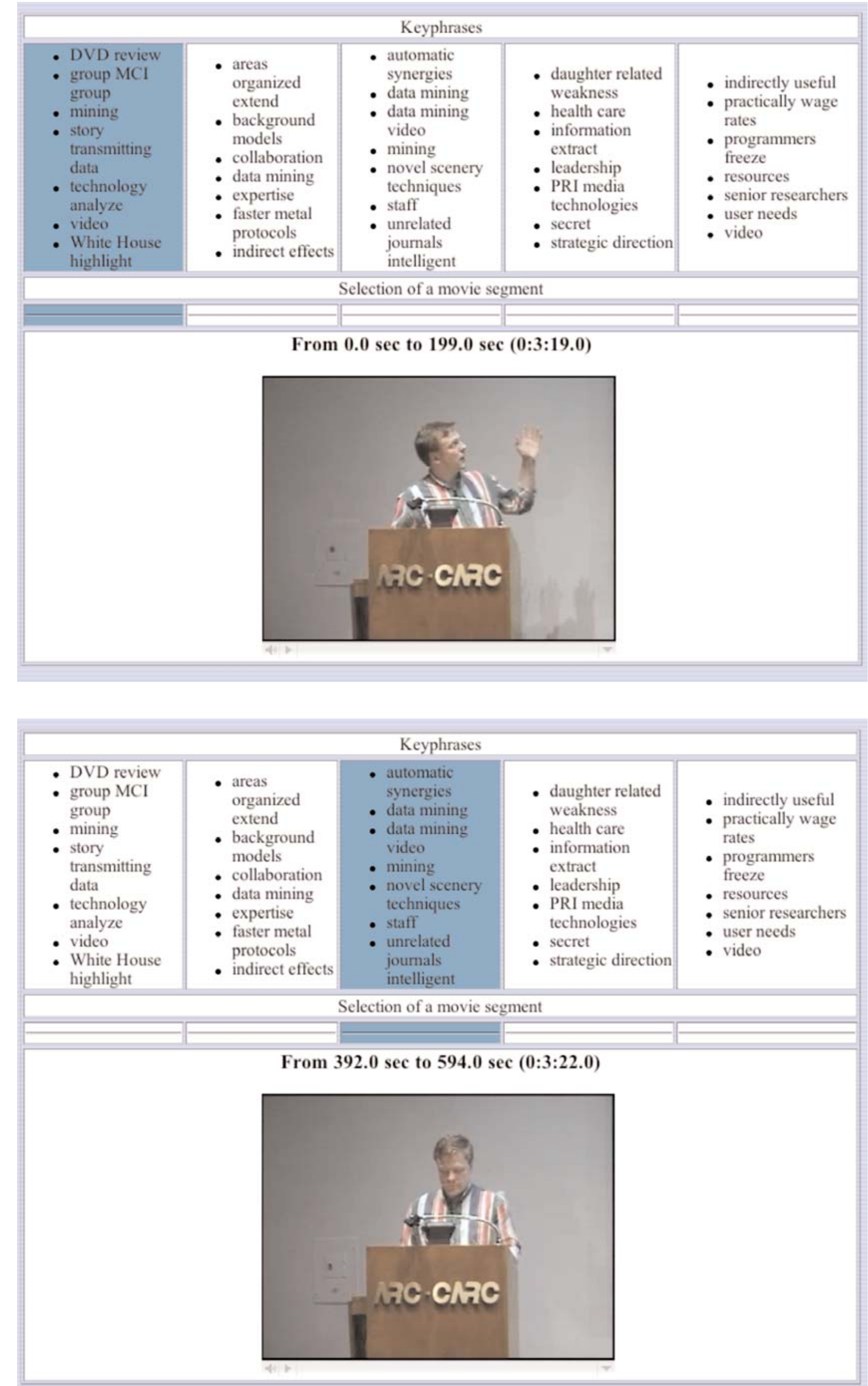

Figure 9: Browsing a video by keyphrases associated with fixed-length video segments.

2.4 Discussion and future research for PVS

Private Video Servers have the potential to rival videoconference in importance among broadband tools for learning. Uses mentioned earlier in this report include:

- Allow teachers to asynchronously share, discuss and reflect on behaviours captured in videos of themselves and their class.
- Facilitate distance collaboration by providing visibility of activities at the individual sites in the intervals between synchronous sessions.
- Improve human performance by supporting timely expert feedback.
- Bind virtual communities by providing a common point of reference on community activity.

This potential was not realized in LearnCanada, but significant lessons were learned:

- Teachers need convenient access to PVS. Asynchronous access occurs when the teacher becomes available; in practice this means that access must be in the classroom.
- Teachers were not all ready to use PVS for peer reflection on their own behaviour. Due to the network issues limiting PVS usability, LearnCanada did not create an environment sufficiently encouraging for teachers to try peer reflection.
- To be useful as a source of feedback, video must be made available immediately following its creation.

Further evidence of the potential for PVS was gathered from several sources:

- Discussion with partners in Finland and Australia indicated that time zones are a limiting factor for their ability to collaborate by videoconference, and that they would welcome PVS as a complement.
- At the National Arts Centre's Young Artist Program in June 2002, we simulated PVS functionality in order to gauge user response. For a week, LearnCanada personnel captured every violin, viola, cello and piano lesson and masterclass, giving the videotape to each student immediately following the lesson. A room with a video player was made available for viewing. We observed students' keen interest in obtaining the video; typical behaviour was to view immediately or soon after the lesson. At the end of the week, a plenary discussion was held with the students, where they described ways in which video could be useful to them.

PVS is an integral component of the MusicGrid[14]. In that context we have separated operational and research activities. Private Video will be served to MusicGrid participants from CRC, using the OptiBase server acquired in LearnCanada. PVS research will continue at NRC, using components more appropriate to research and prototyping. Research results will be trialed with MusicGrid participants in a context where performance expectations are appropriate to evaluation of research.

There are several relevant research areas for PVS:

- Advanced media browsing and search capabilities.
  The purpose of this research is to improve browsing and search capabilities of PVS by using Extractor for Speech to extract keyphrases from the video sound track and use these keyphrases as pointers to video segments.
- Enhanced media player.
  The purpose of this research area is to develop a client application offering more control over the viewing of streamed media, including rapid skimming, fast forward/backward capabilities, as well as a bookmarking functionality for annotation.
- Integration of COTS discussion forums to support discussion about streamed media elements.
  The purpose of this research area is to develop interface tools to support easy integration of PVS with off-the-shelf discussion forum software. This research would fulfil two objectives: augment the annotation capabilities of VSA; and offer to Web-based discussion forum developers a solution to integrate video into their products.
- Integration of remotely distributed video streaming servers.
  The purpose of this research is to evaluate the potential of remotely distributed video streaming servers to support videoconference applications. This server architecture would allow capture and immediate streaming of video content.

3. Videoconference Extension through Webcast (VEW)

The success of videoconferencing learning sessions in LearnCanada and associated projects led to the idea of extending videoconferencing to peripheral participants using webcast technology. Two prototypes of "Videoconference Extension through Webcast" (VEW – pronounced "view") were developed and tested, building directly on the streaming architecture of PVS-2.

3.1 User Functionality

VEW allows people to observe a two-site videoconference. Furthermore, it allows all those observing to communicate with one another, i.e. to discuss the contents of the videoconference. By having a person at a videoconference site monitor the webcast, communication back to the videoconference is established. In effect, this creates a penumbra of partially-participating people around the focal area of people fully participating in the videoconference. This can greatly increase the reach of videoconference.

[14] Proposal currently in submission

3.2 Technology

The solution developed in VEW is to simultaneously digitize and webcast the two video streams of the videoconference. The core of the solution uses the same two-server configuration found in VSA, with additional hardware and software connected to a videoconference unit.

Two versions, VEW-1 and VEW-2, comprised the following components:

Hardware	Network	Software	Standards
• 400MHz Macintosh G3 • MacOS 9.2 • 400MHz Macintosh G4 • MacOS 10.0 • Macintosh G4 Powerbooks • MacOS 9.2	CA*net 3	• cl-http • original code in Macintosh Common Lisp • QuickTime Streaming Server • VEW-1: Sorensen Broadcaster (H263) • VEW-2: QuickTime broadcaster (MPEG4)	• XHTML 1.0[15] • CSS 1.0[16] • SMIL 1.0

Table 3: Technology components in VEW-1

Analog incoming and outgoing video streams from a Tandberg 6000 videoconference unit were converted to digital video (DV) using a Sony DV camera and a JVC DV deck. These devices also captured the streams on DV tape. Firewire (IEEE 1394) fed the two DV streams to two Apple G4 PowerBooks running Sorenson Broadcaster, which encoded the video streams as H263 and encoded the sound with QDesign Music Codec. Mixed sound from the near and far ends was taken as a mono signal from the Tandberg and fed directly to one of the PowerBooks, so that one video stream had both streams' sound; the other video stream had no sound. These independent streams were unicast to the QuickTime Steaming Server, and then brought together as a pair of embedded videos in a webpage.

3.3 Field trials

VEW-1 was tested using a Tandberg 6000 (H320/H323) during violin lessons taught by Pinchas Zukerman in Ottawa with students at the Manhattan School of Music in New York City. The webcast was viewed

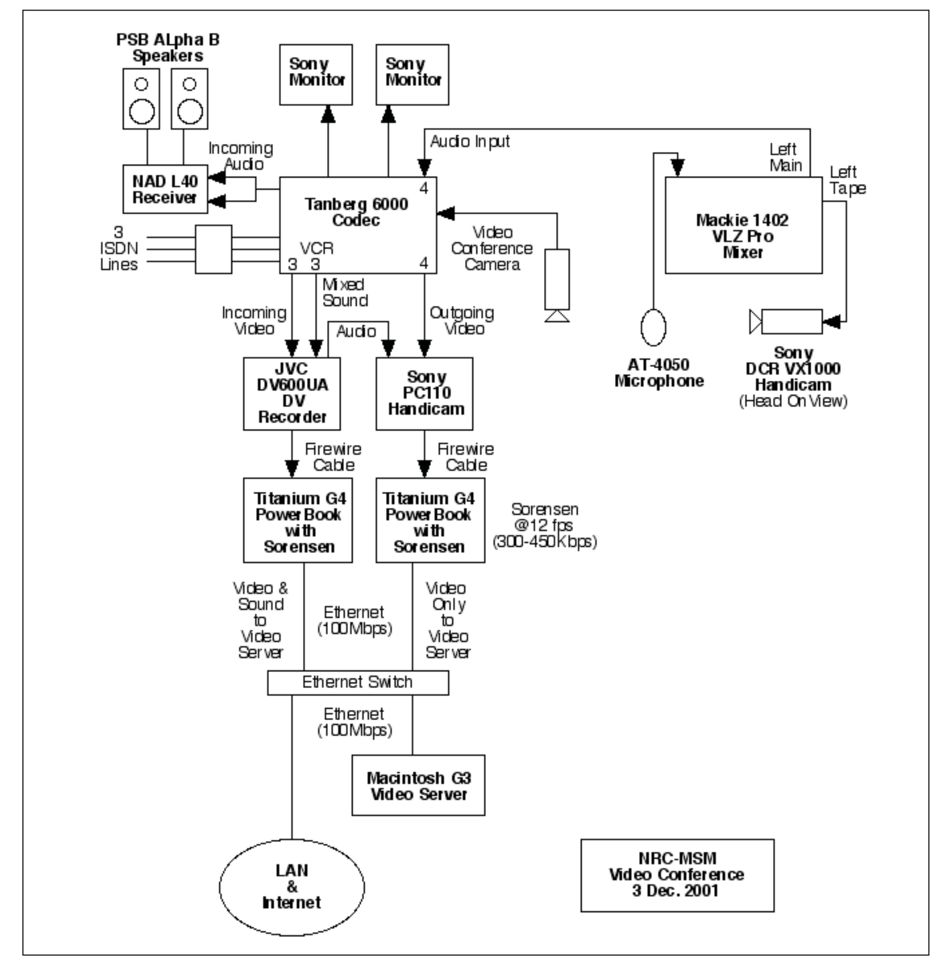

Figure 10: System diagram for VEW-1.

[15] Extensible HyperText Markup Language **(http://www.w3.org/TR/xhtml1/)**

[16] Cascading style sheets **(http://www.w3.org/Style/CSS/)**

only within NRC's building M50. The two independent webcast streams were subject to a lack of synchronization, up to two seconds. Since the mixed sound was associated with a single stream, the sound did not become desynchronized. However, sound quality varied from bad to acceptable, for reasons not yet understood.

Figure 11. VEW-1 in use. Leftmost on the table: webcast client with two video streams.

VEW-2 was tested at the National Art Centre's 2002 Young Artist Program in a test configuration not using videoconference, but instead simulated using two camera inputs. Masterclasses by Pinchas Zukerman and Amanda Forsyth at the University of Ottwa were captured from the back of the auditorium, with one camera tight on the teacher, the other tight on the student. Audio from the mixing board was sent with one video stream. QuickTime Broadcaster generated MPEG4 at on two PowerBooks, each of which unicast to the QuickTime Streaming Server at NRC. A group of colleagues across Canada were invited to observe on CA*net 3. Results were similar to the VEW-1 tests at NRC, with synchronization and sound quality problems.

4. Videoconference Capture

Videoconference recordings can be an important source of information for a virtual community. By providing such recordings on a Private Video Server, they can be reviewed by people not present at the conference, or used as a source of reflection for those that were. They can be used for communication, e.g. in presentations. Videoconference recordings can also be used to study quality factors. LearnCanada developed extensive experience capturing videoconference sessions.

Initially, for Isabel multipoint sessions, a digital video camera was pointed at a projector screen showing all sites, and the camera's microphone picked up the room sound. This was improved by pointing the camera at a plasma screen, and feeding sound directly to the camera from the mixing board. For Isabel sessions, this remains the most practical solution.

When multisite videoconference includes breakout sessions, then the breakout sessions must be captured separately.

Although the individual streams in an Isabel session are not available for separate capture, the two streams in H320/H323 conferences are. They can be captured from the codec, and combined either at production time, e.g. using FinalCut Pro, or at play time, using SMIL.

Figure 12. Screen view of LearnCanada Isabel videoconference session.

Figure 13: Breakout session for sixth graders in VirtualClassroom session.

Figure 14: Captured videoconference (Ottawa – NY); video from two sites coordinated using Synchronized Media Integration Language (SMIL).


```
<smil xmlns:qt="http://www.apple.com/quicktime/resources/smilextensions" >
  <head>
    <layout>
      <root-layout background-color="black" height="241" width="642" />
      <region id="video1" top="1" left="1" height="240" width="320" />
      <region id="video2" top="1" left="321" height="240" width="320" />
    </layout>
  </head>
  <body>
    <par>
      <video region="video1" src="daniel/lalo/3/12/ottawa.mov" begin="0s" end="37s"
type="video/quicktime" />
      <video region="video2" src="daniel/lalo/3/12/nyc.mov" begin="3s" end="37s"
type="video/quicktime" />
    </par>
  </body>
</smil>
```

Figure 15: SMIL script for Figure 14.

Another alternative is that the source video at each site can be separately captured, and then brought together in post-production, using either SMIL or, say, FinalCut Pro. If digital video, with high-quality local sound, is used at each site, this will give the highest quality capture. However, there are synchronization issues that are not easily overcome.

Sound from the other sites must be present in each local recording, or there will be no common reference for synchronization. This is usually not a problem, since far end audio usually leaks into the local microphones. If this were not the case, one could record local audio from the mixer on one channel, and far-end sound on the other. A more ideal solution might be global time from a GPS satellite.

The presence of a common reference does not solve the synchronization problem. Suppose the one-way latency each way is some time d > 0; this is typically several hundred milliseconds. Then video captured at time T at site A is a response to what the people at site A are seeing on their screen from site B, which has taken time d to reach site A. For example, the local sound at site B at time T-d is what site A hears at time T. This suggests that in order to synchronize locally-captured video tapes, one should align tape B's sound at time T-d with tape A's sound at time T. This would indeed show what was experienced at site A. But it is not what was experience at site B; for that one would align tape A's sound at time T-d with tape B's sound at time T.

This choice can be viewed as a positive feature – from the two tapes one can create two videoconference summaries, one showing what was experienced at each end. This would be indeed be useful for studying the effects of latency. For example, in musical sessions it was not uncommon for the teacher to clap in time with the student's playing. Unfortunately, the clapping was not in phase with the musical beats at the student's end.

There is another practical difficulty with this solution: One wants the combined videoconference streams to have the best quality sound, which means using the local sound from each end. But this generally alternates back and forth with the dialog. Production work that cuts each sound track in and out can be very tedious.

5. Publications, Conferences and Multimedia

Emond, B., Brooks, M., Smith, A. (2001). A broadband Web-based Application for Video Sharing and Annotation, In Proceedings of Ninth International ACM Conference on Multimedia, pp. 603-604.

Désilets, A., de Bruijn, B., and Martin, J. (2001). Extracting keyphrases from spoken audio documents. In Information Retrieval Techniques for Speech Applications. A.R. Coden, E.W. Brown, S. Srinivasan (Eds.), pp 36-50.

Barfurth, M.A., Singer, J., Emond, B., Vinson, N., Brooks, M., Spence, J. (2002). Evaluation factors for multi-stakeholder broadband visual communication projects. Proceedings of the Eleventh IEEE workshops on Enabling Technologies: Infrastrcutures for Collaborative Entreprises, Carnegie Mellon University, Pittsburg, PA, June 10-12.

Brooks, M., Emond, B., Spence, J. (2002). Next generation videoconference and video servers for broadband collaboration and learning environments. International symposium on educational conferencing, May 30th – June 1st, Banff, Canada.

See also the list of presentations in the Virtual Community Workpackage report. Most presentations by Brooks and/or Emond explained PVS, videoconference webcast, and/or videoconference capture.

A significant volume of multimedia content was generated during LearnCanada. This includes teacher professional development videos produced at the school boards, and captured videoconferences. The videos have been used as content for discussion in videoconference, as content on the Private Video Server, and for communication about the project.

For example, at Earl of March school in Ottawa, the following videos were created:

1) Jigsaw (2:18) Politics students involved in jigsaw during LearnCanada classroom visitation.
2) EofM Politics (grade 12) students teaching Rideau Civics (grade 10) students (1:24).
3) Jazz Choir (2:20) EofM Jazz Choir during demo to broadband task in St.John's.
4) 4MAT (2:20) grade 12 Politics students at Earl of March learning about the 4MAT model to lesson planning then being involved themselves in the planning of course units.
5) Changing Borders (1:04) Students at EofM discussing post Sept.11th world with JPP and Alberta Justice Minister.
6) LearnCanada Wrap-up session (3:04) recap of Earl's involvment.
7) Visit to the RadioCanada newsroom to better understand the support the newsroom provides the reader (1:15). This is a model we tried to explore as a classroom when using videoconference.

Captured videoconferences include:

- Twenty-three LearnCanada videoconferences;
- Three VirtualClassroom videoconferences;
- Dozens of music teaching videoconferences with Pinchas Zukerman.