Project PATRON—Audio and Video On Demand at the University of Surrey

by Jon Maslin and Elizabeth Lyon

In the March 1997 issue, Derek Law described the "Follet Program" which has provided funding and guided the development of digital library initiatives in the United Kingdom for several years now. In this column Elizabeth Lyon and Jon Maslin describe work at the University of Surrey which is a practical illustration of the success of the program in fostering leading edge experimentation resulting in production systems using a mix of off-the-shelf technology and local development. Project PATRON is an excellent example of the ways that libraries may support the teaching-learning process using technology enabled multi-media—CBL, University of Maryland, College Park, Maryland.

Project PATRON (Performing Arts Teaching Resources ON-line) has been designed to deliver digital audio, video, music scores and dance notation across a high speed network to the desktop. The project was funded through the Higher Education Funding Council for England Joint Information Systems Committee Electronic Libraries Programme and is based in the Library at the University of Surrey, United Kingdom. The principal aim is to investigate ways of improving access to reserve materials, such as music CDs, notation and scores, and dance videos for staff and students. User requirements were investigated through a series of initial focus groups which informed the design of the PATRON interface. User evaluation has continued to play a major part in the project.

The project objectives have been to improve access in terms of multiple users, time, location, browsing, and searching. Many students need to access the same material within a limited period, causing problems in the library. They may also want to have access at the nearest convenient location at all times of the day. In addition, the nature of the subject matter often requires different media to be brought together. For example, a music score may need to be read in conjunction with different recordings.

The Resource

Pilot resource material was selected in conjunction with the School of Performing Arts. The appropriate agreements were obtained for use on the university campus for educational purposes from rights owners and bodies such as the Mechanical Copyright Protection Society, Performing Right Society, and Music Publishers Association. The resource consisted of three types of media: scanned pages of music scores and dance notation, audio, and video.

Conversion

Maintenance of quality was of prime importance because students were to conduct the critical work. For instance, music needed to be of CD quality and the quality of movement in dance video must reflect the original performance closely.

An equally important requirement was that international standards should be adopted for all formats. This meant that, regardless of the hardware platforms and the means of delivery, the material could be accessed, and that there would be an open development path to accommodate future changes in information technology. To an extent, it was also necessary to cope with archival and preservation considerations: even if a standard was superseded, conversion processes would be available.

The images of music scores and dance notation have created particularly demanding problems:

- Originals vary in quality and size;
- Music scores may be reprints, so the detail and contrast can be poor;
- Dance notation was often hand-drawn and could be faint;
- Both types can have fine lines and detail which do not display well on a computer monitor; and
- The application demanded that images display quickly, so files had to be small but include sufficient detail to enlarge without a disturbing loss of quality. The aim was to be able to create the pages as quickly as it takes to turn a page.

Scanned Images

Scanned images had to meet three main criteria: quality, speed, and, for reasons of economical conversion, minimal conversion effort. The need for quality and speed resulted in a compromise since the better the quality, the larger the file size and the slower the building of the image on the screen. However beyond the resolution of the monitor there was no point in increasing the detail of the image, so a determining factor was the acceptability of the print quality, which was effectively a minimum of 300 dpi.

The effect of the resolution and dot size of the monitor was significant because at small magnifications the horizontal stave lines are finer than the dot size, and tend to drop out in oblique bands up the page.

Elizabeth Lyon is Head of Research & Multimedia Development, George Edwards Library, University of Surrey; Jon Maslin is PATRON Project Manager, Information Services, University of Surrey, George Edwards Library, Guildford Surrey, England GU2 5XH.

The method adopted to counter the dot size problem and the quality of the originals was to scan as a grey scale so that the full range of an image could be maintained. Tests were made on each publication to establish threshold levels for white and dark areas. This meant that faint image areas could be maintained or enhanced. Nothing was added to or taken away from the images, so the pages were facsimiles. In order to maintain an economical production operation, pages did not receive individual attention unless they had marks created by users, which were either erased before scanning or by cleaning up in the image processing software.

After scanning, images were reduced in size for displaying on the monitor. The higher quality versions were kept for printing.

Audio

Audio conversion was undertaken in two stages, transfer to a wave format, and then compression. Data compression was important to reduce disk storage requirements and network transmission loads and times. Without data reduction, digital audio signals typically consist of 16 bit samples recorded at a sampling rate more than twice the actual audio bandwidth (e.g., 44.1 KHz for compact disks). So, one second of stereo music in CD quality requires more than 1400 bits.

The international ISO/IEC standard, MPEG, has several levels of compression. The greatest compression is achieved with layer 3, which can compress by a factor of 12, without losing significant sound quality. The compression is achieved largely by elimination of sound outside the perception levels of humans. A number of encoders and decoders are available, for example from the Fraunhofer Institute. Compression decoding is performed in software without the need for special hardware.

Quality was the most important factor in selecting the level of compression. Because staff and students listen to music for a variety of purposes, the decision was taken to adopt the highest perceived quality. A series of tests were made to establish what was acceptable. Subjective tests were conducted on samples of the proposed music resource, and also on sound quality assessment material.

These assessment tracks enabled critical artifacts to be identified. It was apparent to trained ears that the compression algorithm created a very small amount of fuzziness at the beginning and end of sharp sounds. This is not normally apparent, and is unlikely to be distinguished in most applications. While a bit rate of 128 kbits/s would be acceptable for most pieces, it was decided if possible to adopt the highest standard offered by the conversion software, that is, a sampling frequency of 44 KHz, and a bit rate of 256 kbits/s. This resulted in a compression in the region of 1:5.5. However, it was found that the original player/decoder we used was reliable only at 128 Kbits/s, and it was this level which was finally adopted. Later manufactured players do not have this problem.

The conversion operation ran as a series of batch processes, and once set up required minimal attention.

Video

The quality objectives for video were to achieve good broadcast quality because it was important that dance was presented as authentically as possible. Again, MPEG compression was selected. A number of software options were tested, including Optibase Forge, AVID, and the conversion software supplied with the Matrox Studio Runner video card. They all showed a degree of image degradation. The quality of the source material was important; generally Betamax gave better results than VHS.

MPEG1 was used and can be played on a PC with a standard video card, although a high quality video card is preferable. MPEG2 provides higher quality and higher hardware standards. Again, there is a compromise between size of the file and quality.

Tests were conducted to assess the appropriate quality levels and established that with MPEG1 it was necessary to use the top quality, that is, 3.5 Mb/s, and that it may be necessary to use MPEG2. A video with a playing time of 13.5 minutes produced a file of 436 Mb at 3.5 Mb/s, and an MPEG2 conversion at 5 Mb/s produced a file of 637 Mb.

System Components

The library resource was maintained on a file server which for the pilot had a 24 GB of disk storage based on an IBM RS6000 G40. This configuration uses the IBM Multi-media filing system to control access and flow of audio and video files to ensure that uninterrupted streams of data were received by clients. The IBM Digital Library system which provides a range of retrieval and rights and information management functions was also adopted for use.

PATRON is based on the successful transmission of simultaneous streams of multi-media data over a broadband network. There must be no loss of image or sound quality through timing problems. To achieve this goal, high speed (which can be supplied through Fast Ethernet, FDDI or standard Ethernet IEEE 802.3 10Base-T or ATM) was required.

Development and initial evaluation has been conducted on PC's. However, the client application runs in a Web browser, thus eliminating the cross-platform issues of running on Windows based PC's and Mac OS. The user interface is a combination of Java, Java Script, Dynamic HTML, ActiveX controls, and a specially developed module to display graphics images at high speed.

To achieve the objective of improving access, the user client computers were to be as simple to set up as possible. The minimum specification can generally met by most currently available multi-media Pentium II PCs running Windows 95 or NT with a good monitor, video card, and network card.

Information Structure and Access

The end-user's access has to be at very fine detail in the structure of PATRON's information hierarchy. For analytical purposes it has been necessary to access video and audio to a fraction of a second, and to specific places on a page. At higher levels users need to be able to move to sections such as movements or scenes, and at even higher levels they need to access the metadata to find titles, performances, or other descriptive elements.

This has been achieved by three methods which come into play at different stages in the product cycle, and which reflect the economics of creating the appropriate data. Metadata are captured at the earliest stage and are necessary for the rights permission process. They are maintained throughout the life of the item. The data are subsequently used for retrieval through the library catalog which is available over the Internet, through IBM Digital Library and via a simple PATRON search module.

At the capture stage a map is created of each item in the form of a table of contents. This is a hierarchically structured set of data which has an unlimited number of levels. Typically, it is

Table 1
Access Characteristics

Granularity	Tool	Source/Subject Knowledge/Cost
Complete work	PATRON quick search Library OPAC Digital library Metadata Mandatory for PATRON database Possibly extended as a result of further subject classification	From source material at data capture stage. No subject knowledge. Cost relatively low.
Major parts (e.g., movement, scene, page number, and bar number)	Table of contents Mandatory for PATRON database	From source material at data capture stage. Little or no subject knowledge. Cost is variable according to complexity.
Selected detail (e.g., portion of an image or functional selection of short duration)	PRO-file Created only in response to a specific need Personal and subjective	Subject knowledge essential. Cost potential high; therefore, only done on demand.

the table of contents from a score or CD sleeve notes, but it can be extended to the detail required for a particular piece. For example, a score may have pointers to each movement and the bar numbers. As each page is entered in the map functions, repeats can be marked.

Access to the finest detail is usually as a result of expert and subjective analysis which is undertaken after loading, and is likely to continue throughout the life of the resource item. Tools have been created to enable users to record a pointer to the exact position in an audio stream or a page and to place it in a separate document which can be private or public called a *profile*. Table 1 identifies the main characteristics.

User Interface

The user interface had to have a number of attributes: simplicity and robustness; management of the screen territory; search and retrieval mechanisms; and navigation and playing functions.

The user interface works within a Web browser which is simple for users and reliable. One of the ways to achieve simplicity was to manage the screen territory carefully. The most demanding images were music scores which have very fine elements and require the largest possible area of the screen. Since they tend to be vertical, they need the maximum height of the screen.

Potentially the screen can have an unlimited number of windows with audio, video and images. This could become confusing if many applications were started, so a screen layout was adopted which had a main display area and a number of expandable windows. This provided primarily an area for scores with a number of other windows for playing a recording, viewing an image, and conducting a search and for a history of the current session. As needed, windows could be expanded or diminished to fill the main area. In this way the user could switch between an unlimited number of works and media without having a cluttered screen. While this layout was used initially, other layouts, such as a layout in which two scores can be displayed side by side, are optional.

The user interface developed to meet these requirements operated at three levels:

- A search and retrieval process;
- Playing recordings and viewing notation; and
- Establishing links and associations between different sources and media.

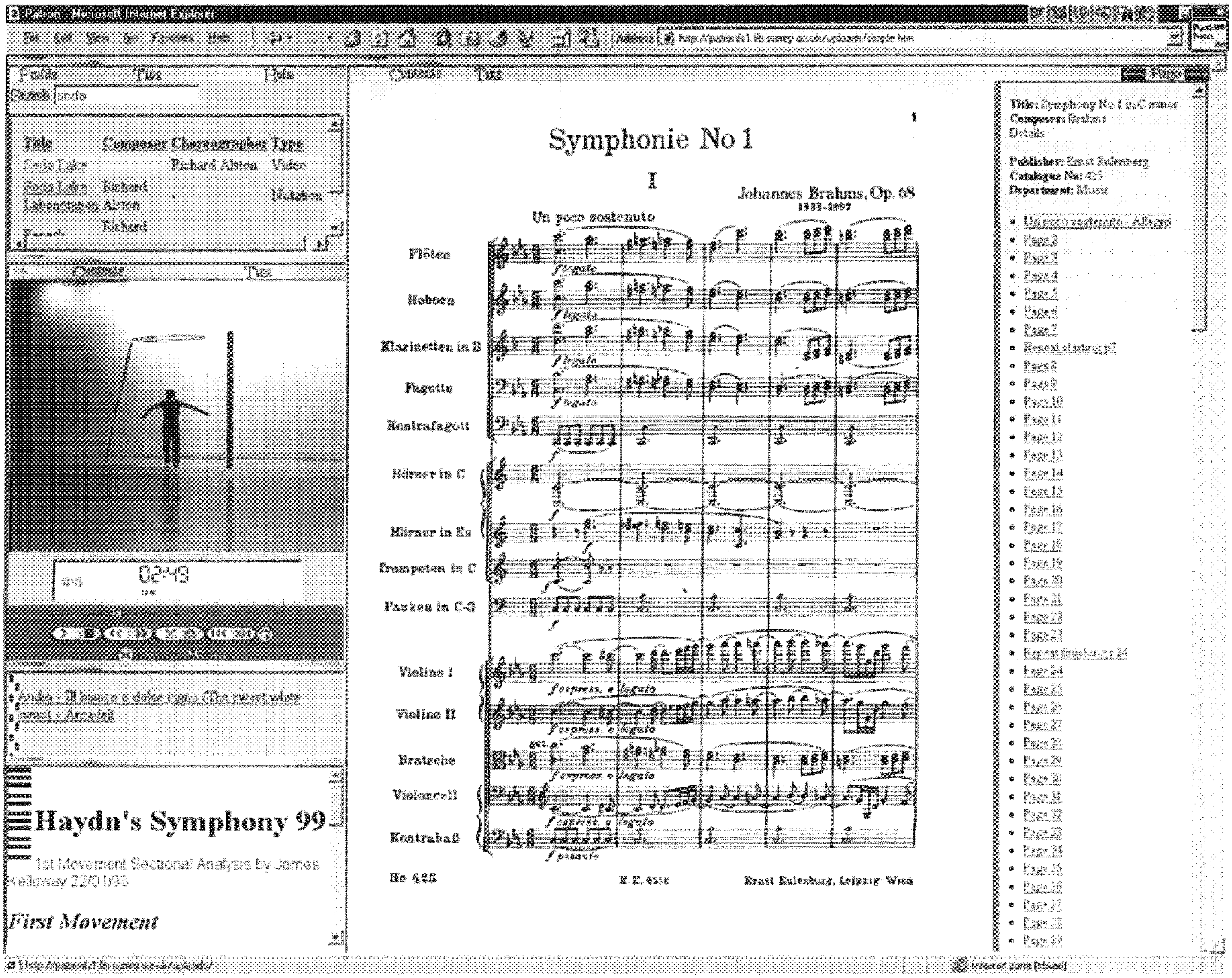
The search and retrieval operation catered to different modes of retrieval. At the simplest was a basic search routine allowing users to enter a string such as "audio mo" which would find all occurrences of audio recordings of Mozart. An alternative provided access through the Web interface to the library catalog, the BLCMP Talis Web OPAC. IBM Digital Library provides opportunities for free-text searching if the information has been entered into the database. In every case, once an item has been identified a single click of a mouse starts it playing.

Within an item, the user had a number of options. The table of contents had links to major points, such as movements or acts. In addition, the table of contents had an expandable list of details about the work, such as conductor, choreographer, and performers.

User functions for reading scores were simple: move to the next page, and go back to the previous page. These mimicked page turning, and were provided for by left and right arrow, or the "n" and "p" keys. Scores were automatically displayed to fill the frame and could be enlarged and dragged around the screen. In order to read the scores it was desirable to have a large screen and to set it at the highest resolution. A 21 inch screen set to 1200x1600 pixels provided good results, but a 17 inch screen set to 1280x1024 was acceptable for much of the time (and was a lot cheaper). A screen setting of 1024x768 was acceptable for pocket book size scores, but with this resolution, zooming was needed for fine detail on large originals.

Audio and video recordings are both played through the same interface. The recording can be played, stopped, started, and replayed or advanced rapidly in ways that are familiar to users of CD or videotape players, but without limitations such as inserting new CDs or tape, serial access, or interference to the picture. In addition there are a number of features which are useful to students: adjusting the speed of playback and making selections which could also be played repeatedly. This was particularly important in the case of video, where a sequence could be played indefinitely without wear to the player or tape. For many users, bringing together a range of items and providing good access to them was sufficient. The full advantage of an electronic resource could be exploited by providing a range of

Figure 1



tools which could be added as they are required. The first has been to develop a means of making links between different items and making annotations about specific points in a piece. This has taken the form of an electronic notebook or document which can be created while the user is listening and viewing. The user interface to achieve this has to be quick and intuitive.

It has been implemented by enabling the user to drag from an icon in the playing frame into the document or pro-file (Patron Resource Organiser File). This automatically dropped in the exact point from a recording or score page. A list of pointers could be organized into groups so that associations may be made between them. Each pointer could be annotated with free-form text. These linkages and comments were maintained independently of the resource so that, for example, an analysis and comment can be made of a score and several interpretations of it.

This pro-file was an HTML document which could be read and edited in any browser or editor. URLs to other Web sites and graphics can be inserted, so once it was created it could be revised, and could be used as, for example, a student assignment, or a lecture cueing list. It is, in essence, a personalized layer which can take many forms, of which this is one. It can also contain programs controlling access, and these have been created in response to a need for students to recognize pieces of music. The program selects random excerpts which can be introduced or summarized by a generated voice from data in the database.

It should be noted that, while a principle use is expected to be made by individuals using headphones, the client can be a computer connected to a high resolution projector and sound amplifier, so enabling lecturers to use library material directly.

Conclusion

The ability to bring together different types of performing arts material through a simple user interface and achieve good levels of quality on standard current computers has been proved. The next stages are to integrate the resource and the associated tools into a number of courses. The focus of the project has been on the performing arts, and it has emphasized the potential value of a digital resource and appropriate user interfaces. It is apparent, however, that the techniques and design approaches developed are valid for any subject matter.

Reactions to the project by all types of potential users have been enthusiastic. At the simplest level, inexperienced computer users have been stimulated by the improvements in accessibility and user controls. Fears which were originally expressed about reading off a screen turned out to be exaggerated, and patrons of PATRON enjoyed using the system. More experienced computer users needed little training, and quickly started to experiment with the pro-file builder. As a new concept, it has required more thought by users, and is expected to be the subject of further development.

The 1998 session of the U.S. Senate and House produced landmark legislation for the future of digital information with respect to the effects of technology on intellectual property and on content itself. In the next column, Prue Adler will unravel the complexities of this legislative watershed and their impact on libraries and scholarly information.

Individuals interested in contributing guest columns should send a précis of their proposed essay to: Charles B. Lowry, Dean of Libraries, Editor, "Managing Technology," JAL, University of Maryland at College Park, 4121 McKeldin Library College Park, MD 20742-7011. Or phone: (301)405-9127; Fax: (301)314-9408; E-mail: "clowry@deans.umd.edu".